

100

SwissProt_40:HX43_HAELIN	198.00	171.89	0.0077	917	P43553 haemophilus influenzae	1125
SwissProt_40:HM44_MOISE	198.00	170.16	0.0078	1125	P27546 mus musculus (mouse)	1162
SwissProt_40:TCNA_TRECR	198.00	169.88	0.0078	1162	P32253 trypanosoma cruzi	1902
SwissProt_40:P2P_LACCA	197.00	164.83	0.0091	1902	Q02470 lactobacillus paracellus	1120
SwissProt_40:TRXB_ECOLI	196.50	168.87	0.0092	1120	P76072 escherichia coli	3726
SwissProt_40:SUBR_BACSU	196.00	158.25	0.0108	3726	P20659 drosophila melanogaster	1433
SwissProt_40:FMG2_YEAST	195.50	165.90	0.0106	1433	P16397 bacillus subtilis	1609
SwissProt_40:APC_RAT	195.00	164.48	0.0113	1609	P25653 saccharomyces cerevisiae	2842
SwissProt_40:HMES_DROME	194.00	158.78	0.0133	2842	P70478 rattus norvegicus (brown rat)	1077
SwissProt_40:YMOB_CAEEL	193.50	166.56	0.0129	1077	Q02308 drosophila melanogaster	918
SwissProt_40:HX41_HAELIN	191.00	165.83	0.0169	918	P34487 caenorhabditis elegans	905
SwissProt_40:VACA_HELPY	191.00	162.82	0.0174	1291	Q48258 haemophilus influenzae	1902
SwissProt_40:P2P_LACCA	189.50	158.21	0.0214	1902	P15293 lactococcus lactis	1113
SwissProt_40:N1P6_YEAST	188.50	161.87	0.0228	1113	Q02630 saccharomyces cerevisiae	2845
SwissProt_40:APC_RAT	188.50	153.92	0.0248	2845	P03135 mus musculus (mouse)	1829
SwissProt_40:PRPC_NETMB	188.00	157.22	0.0252	1829	Q93YV5 neisseria meningitidis	1287
SwissProt_40:YAC2_HELPY	187.00	159.32	0.0274	1287	Q48245 haemophilus influenzae	1045
SwissProt_40:PRPS_SEEWA	186.50	160.64	0.0285	1045	P03489 serraria marcescens	2843
SwissProt_40:APC_HUMAN	185.50	151.28	0.0348	2843	P25054 homo sapiens (human)	1337
SwissProt_40:DEXT_STPDO	185.00	157.23	0.0345	1337	P33653 streptococcus dysenteriae	1480
SwissProt_40:PANI_YEAST	185.00	156.37	0.0348	1480	P35251 saccharomyces cerevisiae	1829
SwissProt_40:FRPC_NETMC	185.00	154.58	0.0354	1829	P25127 neisseria meningitidis	1902
SwissProt_40:P3P_LACCA	185.00	154.24	0.0355	1902	P15392 lactococcus lactis	1419
SwissProt_40:ALAI_CANAL	184.00	155.84	0.0388	1419	Q13368 candida albicans	660
SwissProt_40:THLI_EBY	181.50	160.12	0.0482	660	P03181 epstein-barr virus	1902
SwissProt_40:PIB_LACCA	181.50	151.16	0.0528	1902	P16271 lactococcus lactis	1273
SwissProt_40:WBI_YEAST	181.00	154.12	0.0539	1273	P33658 saccharomyces cerevisiae	986
SwissProt_40:YHDP_ECOLI	180.00	155.40	0.0591	986	P46474 escherichia coli	1520
SwissProt_40:PMPP_CHLMU	180.00	151.73	0.0613	1520	Q96180 chlamydia muridarum	1753
SwissProt_40:YEM2_YEAST	180.00	150.53	0.0631	1753	P33634 saccharomyces cerevisiae	1616
SwissProt_40:SLAP_BACCI	179.50	150.77	0.0653	1616	P33824 bacillus circulans	2124
SwissProt_40:PECA_RAT	179.50	148.46	0.0668	2124	P07897 rattus norvegicus	790
SwissProt_40:NMP_NOTO	178.50	155.95	0.0687	790	P24856 notochorda coriacea	1211
SwissProt_40:PMPP_CHLMU	178.50	154.07	0.0700	987	Q92445 chlamydia muridarum	1296
SwissProt_40:BUN2_DROME	178.50	152.34	0.0713	1211	Q24523 drosophila melanogaster	797
SwissProt_40:VAC1_HELPY	178.50	151.76	0.0717	1296	Q48247 haemophilus influenzae	1130
SwissProt_40:VGIX_HAEB	177.50	155.00	0.0770	797	P28968 equine herpesvirus 1	936
SwissProt_40:N121_RAT	177.50	151.54	0.0797	1130	P52591 rattus norvegicus	1130
SwissProt_40:REPT_MOOSE	177.00	151.60	0.0839	1130	P93747 mus musculus	1288
SwissProt_40:RAE_CITR	176.50	152.75	0.0874	936	Q07551 citrobacter freundii	2090
SwissProt_40:VACA_HELPY	176.50	150.05	0.0898	1288	Q92K45 haemophilus influenzae	1754
SwissProt_40:N121_HUMAN	176.50	145.99	0.0937	2090	P33658 homo sapiens (human)	1754
SwissProt_40:PMPP_CHLMU	176.50	146.99	0.0976	1754	Q84418 chlamydia trachomatis	2201
SwissProt_40:TEMA_HUMAN	175.00	145.07	0.0996	2201	P24821 homo sapiens (human)	2175
SwissProt_40:HMCO_DROME	175.00	144.73	0.1052	2175	P10180 drosophila melanogaster	973
SwissProt_40:PM13_CHLPH	174.50	150.66	0.1099	973	Q92896 chlamydia pneumoniae	1014
SwissProt_40:HEX2_YEAST	174.50	150.31	0.1103	1014	Q00816 saccharomyces cerevisiae	1026
SwissProt_40:VG37_BP14	174.50	150.21	0.1105	1026	P03744 bacteriophage t4	1229
SwissProt_40:N121_HUMAN	174.00	148.24	0.1187	1229	Q97243 homo sapiens (human)	1238
SwissProt_40:SBC1_HOCCA	173.50	147.74	0.1287	1238	Q68032 rhodococcus capsulatus	1243
SwissProt_40:VG37_BP13	172.00	146.38	0.1490	1243	Q38394 bacteriophage k3	1460
SwissProt_40:PM24_CHLMU	172.00	145.02	0.1511	1460	Q93YJ1 chlamydia muridarum	940
SwissProt_40:WACA_SCHCO	171.00	147.87	0.1629	940	P37338 schizopbyllum commune	1597
SwissProt_40:STF1_STPDO	171.00	143.38	0.1705	1597	P11001 streptococcus downii	2426
SwissProt_40:CON1_HUMAN	171.00	139.84	0.1767	2426	P18583 homo sapiens (human)	437
SwissProt_40:YGY3_HAISO	170.50	153.91	0.1613	437	P21561 haloferax sp. (strain)	1140
SwissProt_40:YV96_YEAST	170.50	145.79	0.1752	1140	Q04893 saccharomyces cerevisiae	
seq_name: SwissProt_40:IGAL_HAELIN						
seq_documentation_block:						
ID	IGAL_HAELIN	STANDARD:	PR7:	1541	AA.	
AC	P42782:					
DT	01-NOV-1995	(Rel. 32, Created)				
DT	01-NOV-1995	(Rel. 32, Last sequence update)				
DT	01-NOV-1995	(Rel. 32, Last annotation update)				
DE	Immunoglobulin A1 protease precursor (EC 3.4.21.72) (IGAL protease).					
GN	IGA.					
OS	Haemophilus influenzae.					
OC	Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;					
NC	Haemophilus					
NC	taxid=727;					


```
851 ..... 851
2604 TGAACAGCCGCTTACCGGAAATTCAGCGCGGCAAGATACGGCAT 2653
851 ..... 851
2654 TACACTTAAAGACGGAATGAGCTCCGTCGGGACGGAATTAGCC 2703
852 .....SerHisThrHisLeuThrGlyAsnSerAspValHis 863
2704 AATTAACTGTACACGCCACATTCATCATTCCGCTATCGACA 2753
864 GlnLeuAspLeuAlaAsnGlnHisLeuHisLeuSerAlaAspAsnSe 880
2754 CGATGGGACGAGCGGCAACCGGACGTGCGAGATGCGCGCGCGCC 2803
880 rAsn..... 881
2804 GTTCGGCGCTCCCTATTATCCGTTACGCCGCAACTTCGGCAGAAATCC 2853
882 .....AsnValThr..... 884
2854 CGTTTACACGCTGACGTTAAACGCAATTCGAACGCTCAGGACATC 2903
885 LysTyrAsnThrLeuThrValAsn...SerLeuSerGlyAsnGlySerPh 900
2904 CCGCTTTATGTGGAACCTCTCGGCTACCGGACGCGCAATTCGAAGCTGG 2953
900 eTyTyTyLeuThrAspLeuSerAsnLysGlnLysAspLysValValT 917
2954 CGGAAGTTCCGAGGCACTTACCTTGGCTGTCAACATACCGGCAAC 3003
917 hTySerAlaThrGlyAsnPhetThrLeuGlnValAlaAspLysThrGly 933
3004 GAACCCGTAAGTCTGAGCAATTCGATGAGTGAAGGAAAGAACACAC 3053
934 GluPro...AsnHisAsnGlnLeuThrLeuPheAspAla.....SerLys 947
3054 ACCGCTGTCCGAAATCTTAATTCACCTCGCAAAACGAACAGTCGATG 3103
947 salaglnatrgasphHisLeuAsnValSerLeuValGlyAsnThrValAspL 964
3104 CCGGCGCATGGCGTTATCAGCTTATCCGCAAGACGCGGATTCGCGCTG 3153
964 euGlyAlaTrpLysTyTyLeuArgAsnValAsnGlyArgTyAspLeu 980
3154 CATAATCCG..... 3162
981 TyrAsnProGlnValGlnLysArgAsnGlnThrValAspThrAsnIl 997
3162 ..... 3162
997 eThrThrProAsnAsnIleGlnAlaAspValProSerValProSerAsnA 1014
3162 ..... 3162
1014 snGlnGlnIleAlaArgValAspGlnAlaProValProProProAlaPro 1030
3162 ..... 3162
1031 AlaThrProSerGlnThrThrGlnThrValAlaGlnAsnSerLysGlnI 1047
3162 ..... 3162
1047 userLysThrValGlnLysAsnGlnAspAlaThrGlnThrThrAlaG 1064
3163 .....GTCAAGACACAAGAG 3177
1064 lAsnArgGlnValAlaLysGlnAlaLysSerAsnValLysAlaAsnThr 1080
3178 CTTTCGACAAACTCGGCAAGCGGGA...GAACAGAGCGCGCTTGAC 3224
1081 GlnThrAsnGlnValAlaGlnSerGlySerGlnThrLysGlnThrGlnTh 1097
3225 GGCAAAACAGGCACAACTTGCCGCCCAAAACAAACAGCGGCAAAACAGCAACG 3274
1097 rThrGlnThrLysGlnThrAlaThrValGlnLysGlnLysAlaLysV 1114
3275 CGCAAAAGC.....CTTGACGCCCTGATTGCG 3300
1114 alGlnThrGlnLysThrGlnGlnValProLysValThrSerGlnValSer 1130
3301 GCCGGGCGCAATGCCACCGCAAAAGCAGAAAGTTCGCCAACCGCGCCG 3350
1131 ProLysGlnGlnLysSerGlnThrValGlnProGlnAlaGlnProAlaAr 1147
3351 CGAGCGACGCGGGAATGCGGCAATTATGACGCGGAGAGAGAGAAA 3400
1147 g.....GlnAsnAspProThrValAsnIleLysGlnProGlns 1160
3401 AACGGGTGCAGCGCGATTAACACACCGCCTTG...GCGAAACAGCGGAA 3447
1160 eGlnThrAsnThrThrAlaAspThrGlnProAlaLysGlnThrSer 1176
3448 GCG..... 3450
1177 SerAsnValGlnGlnProValThrGlnSerThrThrValAsnThrGlyAs 1193
3451 .....GAACCGCGCGGCTACACCGCCTTCC 3478
1193 nSerValValGlnAsnProGlnAsnThrThrProAlaThrThrGln...P 1209
3479 CC.....CGCGCCCGCGCGC 3492
1209 roThrValAsnSerGlnSerSerAsnLysProLysAsnAlaGlnHisArg 1225
3493 GCCCGCGGGAATTGCGG...CAACCGGACGCCCAACCGCAACCCCAAC 3539
1226 SerValArgSerValProHisAsnValGlnProAlaThrThrSerAs 1242
3540 GCAGCGC.....GACCTGATCAGCCGTTATGCCAATA 3571
1242 nAspArgSerThrValAlaLeuCysAspLeuThrSerThrAsnThrAsnA 1259
3572 GCGGTTGAGTGAAATTTTCGCC.....ACGCTCAACAGC 3606
1259 lValIleSerAspAlaArgAlaLysAlaGlnPheValAlaLeuAsnVal 1275
3607 GTTTTCGGCGTACGAGCAATTGACCGGCTG...TTGCCGAGACGCG 3653
1276 GlnLysAlaValSerGlnHisIleSerGlnLeuGlnLysAsnGlnI 1292
3654 CCGCAACGCGCTTTGGACAACGCGCATCGGGACCAACCACTACCGTT 3703
1292 yGlnTyAsnValTrpValSerAsnThrSerMetAsnLysAsnTySerS 1309
3704 CGCAAGATTTCCGCGCTACCGCCACAAACCGACCTGCGGCAATCGGT 3753
1309 eSerGlnTyArgArgPheSerSerLysSerThrGlnThrGlnLeuGly 1325
3754 ATGCAGAAAACCTCGGACAGCGG...CGGCTGCGGCACTCGTTTGGCA 3800
1326 TrpAspGlnThrIleSerAsnAsnValGlnLeuGlyValPheThrTy 1342
3801 CAACCGACCGGAAACACTTCGACGACGAGCATCGGACCTCGGACGCGC 3850
1342 rValArgAsnSerAsnAsnPheAspLysAlaThrSerLysAsnThrLeuA 1359
3851 TTGCCCAAGGTGCCGTTTTCGGCAATACGCGCATCGGACGTTGACATC 3900
1359 lAsnValAsnPhetYrSerLysTyTyAlaAspAsnHisTyTyLeu 1375
3901 GGCATCAGCGCGGCGCGGCTTTAGTAGAGCGCACTTCACAGCGCAT 3950
1376 GlyIleAspLeuGlyTyTyGlyLysPheGlnSerLysLeuGlnThrAsnH 1392
```



```

3951 CAGAGCGAAATCCGCCGCGTCTGCATTACGCGATTACGCAAGAT 4000
1392 sasnAlaIysPheAlaArgHisThrAlaGlnPheGlyLeuThrAlaGlyL 1409
4001 ACCGGCAGATTTCGGCGGATTCGCATTCGACCGCAGATCGGCGCAAG 4050
1409 ysalAlaPheAsnLeuGlyAsnPhelGlyLeuThrProLeuValGlyVal 1425
4051 CGCTATTTCGTCGCAAAAGCGATTCACGATACGAAAGCGATTCGCG 4100
1426 TyserTyrLeuSerAsnAlaPheAlaLeuAspAlaAlaArgGlyLeu 1442
4101 CACCCCGCGCGCTTCATTCACCGCTACCGCGCGCGCATTAAGCGAGAT 4150
1442 sValAsnProLeuSerValThrAlaPheAlaGlnValAsnValAsnGly 1459
4151 ATTCATTCACCGCGCGCGCATTCATTCACCGCTTCATTCATTCAGCTG 4200
1459 yThrTyrHisLeuGlyGlu...PheSerValThrProLeuLeuSerAla 1474
4201 TCCTATACGATCCGCTCGCGCAAGAGTCGCAAGCGCGCTCAATACCG 4250
1475 ArgTyr...AspAlaAsnGlnGlySerGlyLysIleAsnValAsnGlyTyr 1490
4251 CGTATTGGCGCAGATTCGCGCAAAACCGCAGTGGCGGCGGCTAA 4300
1490 rAspPheAlaTyrAsnValGluAsnGlnGlnIleThrAlaGlyLeuL 1507
4301 AGCCGCAATCAAGGTTTCACGCTGCTCCGCTACCGCTCGCGCGCGCA 4350
1507 ySerLeuTyrHisAsnValLysLeuSerLeuIleGlyGlyLeuThrLys 1523
4351 GGGCGCGCATTCGCAAGCGCAGCAGCGCGCGCATCAATTCGCGTAC 4398
1524 AlaIysGlnAlaGlnLysGlnIleThrAlaGlnLeuLysLeuSerPhe 1539

seq_name: SwissProt_40:IGAO_HAEIN

seq_documentation_block:
ID IGAO_HAEIN STANDARD; PRT; 1694 AA.
AC P44969;
DT 01-NOV-1995 (Rel. 32, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE Immunoglobulin A1 protease precursor (EC 3.4.21.72) (IGa1 protease).
DE IGA OR IGA1 OR HI0990.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
CX NCBL_Taxid=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-SEROTYPE D;
RA Wright A., Fishman Y., Tai F., Plant A.G.;
RL Submitted (MAR-1991) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-RD / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Klenzwe A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Usterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA Grehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus
RT Influenzae Rd."
RL Science 269:496-512(1995).
CC -!- FUNCTION: VIRULENCE FACTOR; CLAVES HOST IMMUNOGLOBULIN A
CC PRODUCING INTACT FC AND FAB FRAGMENTS.

```

```

CC -!- CATALYTIC ACTIVITY: Cleavage of immunoglobulin A molecules at
CC certain Pro-I xaa bonds in the hinge region. No small molecule
CC substrates are known.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- DOMAIN: THE SIGNAL PEPTIDE GUIDES THE PRECURSOR TO THE PERIPLASMIC
CC SPACE, AND THE CARBOXY-TERMINAL HELPER DOMAIN ASSOCIATES WITH THE
CC OUTER MEMBRANE TO FORM A PORE FOR EXCRETION OF THE PROTEASE
CC DOMAIN. THE HELPER DOMAIN IS THEN RELEASED BY AUTOPROTEOLYSIS (BY
CC SIMILARITY).
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S6 (SERINE PROTEASE).
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: X59800; NOT_ANNOTATED_CDS.
CC EMBL: U32779; AAC22651.1; -.
CC DR MEROPS: S06.001; -.
CC DR TIGR: H10990; -.
CC DR InterPro: IPR000710; IGA_S6.
CC DR Pfam: PF02395; IGA1; 1.
CC DR PRINTS: PR00921; IGASERPTASE.
CC DR Complete proteome.
CC KW Hydrolyase; Serine protease; Transmembrane; Zymogen; Signal;
CC K1
CC FT CHAIN 1 25
CC FT PROPEP 1015 1694
CC FT ACT_SITE 288
CC FT CONFLICT 253 254
CC FT CONFLICT 272 272 G -> A (IN REF. 1).
CC FT CONFLICT 464 464 G -> E (IN REF. 1).
CC FT CONFLICT 866 866 S -> T (IN REF. 1).
CC FT CONFLICT 1036 1036 A -> D (IN REF. 1).
CC FT CONFLICT 1074 1074 A -> G (IN REF. 1).
CC FT CONFLICT 1421 1421 A -> G (IN REF. 1).
CC FT CONFLICT 1545 1545 H -> T (IN REF. 1).
CC SQ SEQUENCE 1694 AA; 185539 MW; C52427013F93178C CRC64;

alignment_scores:
Quality: 1587 50 Length: 1859
Ratio: 1.683 Gaps: 48
Percent Similarity: 50.726 Percent Identity: 25.282

alignment_block:
US-09-303-518D-653 x IGAO_HAEIN ..
Align seg 1/1 to: IGAO_HAEIN from: 1 to: 1694
64 CGCTTCGCGCGCTTACTAGCATATGCTGCTGCGCATTCGCG 113
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
5 LysPheLysLeuAsnPhelAlaLeuThrValAlaTyrValaLeuThrPr 21
114 CCAAGCGCGCGCGGAGACACTTATTTCGCGCATCACTACCAATACTATC 163
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
21 oTyrThrGluAlaValaLeuValArgAspAspValaAspTyrGlnIlePhe 38
164 GCGCATTCGCGCAAAATTAAGCAAGTTTCGCTGCGCGCGCAAGATATT 213
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
38 rAspPheAlaGlnValAsnLysGlyArgPheSerValaLysAlaThrAsnVal 54
214 GAGCTTACCACAAAAAGGAGGAGTGTGCGCAAAATCGATGACGAAGC 263
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
55 GluValArgAspLysAsnAsnHisSerLeuGlyAsnValLeuProAsnG 71
264 C...CGATGATGATTTTCTGTGTCGCG...CGTACGCGCGTGGCG 307
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
71 yLysProMetLysAspPheSerValaValaAspValaLysArgIleAlaT 88

```



```
655 etAsnGlyPheAsnGlyTyrPheGlyGluGluGly...LysAsnAsn 670
1816 GGGCGGCGTAACTGTAATTACCAACGGGAGAGAGCGGATCGCACTTTACT 1865
|||||
671 GlyAsnLeuAsnValThrPheLysGlyLysSerGluGlnAsnArgPheLe 687
1866 GCTTTCGGCGGAGCAAAATTTAAAGGCGAATATCAACGCAACAGCGCA 1915
|||||
687 uLeuThrGlyGlyThrAsnLeuAsnGlyAsnLeuLysValGluLysGlyT 704
1916 AACTGTTTTGAGCGGAGAGCGGAGAGCGGAGCGGATCAATCATTTAGGA 1965
|||||
704 hLeuPheLeuSerGlyArgProThrProHisAlaArgSerPheAlaGly 720
1966 AGCGGCGTGTCAAAAATGGAAGT....ATCCCAACAAGGAAATGCT 2009
|||
721 lIeSerSerThrLysAlaSerPheAlaHisPheAlaGluAsnAsnGluValVa 737
2010 GTGGGACAACGATTTGATCGACGCGACATTTAAAGCGGAAAACTTCATA 2059
|||||
737 lValGluAspAspTrpIleAsnArgAsnPheLysAlaThrAsnIleAsnV 754
2060 TTCAGGCGCGACAGCGGTTGTC...CGAATGTTCCAAAGTGGAA 2106
|||||
754 aLThrAsnAsnAlaThrLeuTyrSerGlyArgAsnValAlaAsnIleThr 770
2107 GCGCATTTGGCATTTAGCAATCAGCGCCAGAGAGTTTCGGTGTGGCACC 2156
|||||
771 SerAsnIleThrAlaSerAspAsnAlaLysValHisIleGlyTyr.... 785
2157 GCATCAAAAGCCACACATCTGTACAGCTTCGAGTGCAGCGGTCGACAA 2206
|||||
786 LysAlaGlyAspThrValLysValArgSerAspTyrThrGlyValT 802
2207 GTTGACGGAATAAACAATTTACCGGAGTAAAGTATGCTTCATTTAGC 2256
|||||
802 hCysThrThrAspLysLeuSer...AspLysAlaLeuAsnSerPheAsn 817
2257 AAGACCGACATCAGAGCAATGTACGCTTCGCGATTCACGCTCAATTTAA 2306
|||||
818 AlaThrAsnValSerGlyAsnValAsnLeuSer..... 828
2307 TCTCACAGGACTTGGCACACTCAAGCGCAATCTTAGTCAGGCGGAGACA 2356
828 ..... 828
2357 CGCACTATACGGTTACGGCAAGCCACCAAAAGCGCAACCTCAGCCTC 2406
828 ..... 828
2407 GTGGGCAATGCCCAAGCAACATTTAATCAAGCCACATTTAAAGCGCAAC 2456
|||||
829 ...GlyAsnAlaAsnPheValLeuGlyLysAlaAsnLeuPheGlyThrI 844
2457 ATCGGCTCGGACATGCTTCATTTAATCAATCAAGCAACAGCGGATCAAA 2506
|||||
844 eSerGlyThrGlyLysSerGlnValArgLeuThrGluAsn..... 857
2507 ACGGCACTGTACGCTTCGCAACAAGCTAAGGCAAAAGTAAGCATTTCC 2556
857 ..... 857
2557 GCACTCACGCAATGTCTCCCTAGCCGATTAAGCAATTTCCATTTTGA 2606
857 ..... 857
2607 AAACAGCGCTTTACCGGAAAAATCAGCGGCGGCAAGATACGGCATTTAC 2656
857 ..... 857
2657 ACTTAAAGACAGCGAATGACGCTCGCGTGGCGACGGAATTAGGCAAT 2706
|||||
858 .....SerHisTrpHisLeuThrGlyAspSerAsnValAsnGln 870
2707 TTAACCTTGACACAGCCACCATTTACACTCAATTCGCGCTATGACACGA 2756
|||||
871 LeuAsnLeuAspLysGlnHisIleLeuAsnAla.....GlnAsnAs 885
2757 TGGCGAGGCGCGCAACCGGCACTGCGGAGATGCGCGCGCGCGCGCTT 2806
|||||
885 pAlaAsnLysValThrThr..... 891
2807 GCGCGCTTCCTATTTATCCGTTACGCGCGCAACTTCGCGAGATCCCGT 2856
891 ..... 891
2857 TTCAACACGCTGACGCGTAAACGCGCAATTTGACGCGTACGGAACATTCG 2906
|||||
892 TyrAsnThrLeuThrValAsn...SerLeuSerGlyAsnGlySerPheTyr 2907
2907 CTTTATGTCGAACTTTCGCTACCGCGACGCGCAATTTGACGCGGCGG 2956
|||||
907 rTyrLeuThrAspLeuSerAsnLysGlnGlyAspLysValValValThrL 924
2957 AAAGTTCGGAAGGCATTTACACCTTGGCTGTCAACAATACCGGCAACGAA 3006
|||||
924 ySerAlaThrGlyAsnPheThrLeuGlnValAlaAspLysThrGlyGlu 940
3007 CCGCTAGTCTCGACCAATTTGACGCGTGTGGAAGGAAACACACACACC 3056
|||||
941 ProThrLys...AsnGluLeuThrLeuPheAspAlaSerAsnAlaThr... 955
3057 GCTGTCCGAAATCTTATTTTCACTCCGCAAAAGCAACACATCGATGCGG 3106
|||||
956 ...ArgAsnAsnLeuAsnValSerLeuValGlyAsnThrValAspLeu 971
3107 GCGATGCGCTTATCAGCTTATCGGCAACAGCGGAGTTCCGCTGCAT 3156
|||||
971 lYalatrPlysTyrLysLeuArgAsnValAsnGlyArgAspLeuTyr 987
3157 AATCCG..... 3162
988 AsnProGluValGluLysArgAsnGlnThrValAspThrThrAsnIleTh 1004
3162 ..... 3162
1004 rThrProAsnAsnIleGlnAlaAspValProSerValProSerAsnAsnG 1021
3162 ..... 3162
1021 lGluIleAlaArgValGluThrProValProProAlaProAlaThr 1037
3162 ..... 3162
1038 ProSerGluThrThrGluThrValAlaGluAsnSerLysGlnLysLeu 1054
3163 .....GTCAAAGACACAGAGCTTTCGCAACAACCTCGCAAGCGG 3202
|||||
1054 sThrValGluLysAsnGluGlnAspAlaThrGluThrThrAlaGlnAsnG 1071
3203 GAGAA..... 3207
1071 lYgluValAlaGluGluAlaLysProSerValLysAlaAsnThrGlnThr 1087
3208 .....ACAGA 3212
1088 AsnGluValAlaGlnSerGlySerGluThrGluGluThrGlnThrThrG 1104
3213 GCGCGCTTGACGCA..... 3228
1104 uIleLysGluThrAlaLysValGluLysGluGluLysAlaLysValGlu 1121
3229 .....AAA 3231
1121 yAspGluIleGlnGluAlaProGlnMetAlaSerGluThrSerProLys 1137
```


OS Haemophilus influenzae.
 OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
 OC Haemophilus.
 OX NCBI_TaxID=727;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=HK393 / NCTC 8467 / SEROTYPE B;
 RX MEDLINE=92234949; PubMed=1373717;
 RA Poulsen K., Reinholdt J., Kilian M.;
 RT "A comparative genetic study of serologically distinct Haemophilus
 influenzae type 1 immunoglobulin A1 proteases."; ,
 RL Bacteriol. 174:2913-2921(1992).
 CC -1- FUNCTION: VIRULENCE FACTOR; CLEAVES HOST IMMUNOGLOBULIN A
 CC PRODUCING INTACT FC AND FAB FRAGMENTS.
 CC -1- CATALYTIC ACTIVITY: Cleavage of immunoglobulin a molecules at
 CC certain Pro-I-Xaa bonds in the hinge region. No small molecule
 CC substrates are known.
 CC -1- SUBCELLULAR LOCATION: Secreted.
 CC -1- DOMAIN: THE SIGNAL PEPTIDE GUIDES THE PRECURSOR TO THE PERIPLASMIC
 CC SPACE, AND THE CARBOXY-TERMINAL HELPER DOMAIN ASSOCIATES WITH THE
 CC OUTER MEMBRANE TO FORM A PORE FOR EXCRETION OF THE PROTEASE
 CC DOMAIN. THE HELPER DOMAIN IS THEN RELEASED BY AUTOPROTEOLYSIS (BY
 CC SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S6 (SERINE PROTEASE).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC DR EMBL; M87490; AAA24967.1; .
 CC DR MEROPS; S06.001; .
 CC DR InterPro; IPR000710; IGA_S6.
 CC DR Pfam; PF02395; IGA1; 1
 CC DR PRINTS; PR00921; IGASERPTASE.
 CC KW Hydroxylase; Serine protease; Transmembrane; Zymogen; Signal.
 CC FT SIGNAL 1 25 POTENTIAL.
 CC FT CHAIN 26 1012 IMMUNOGLOBULIN A1 PROTEASE.
 CC FT PROPEP 1013 1545 HELPER PEPTIDE (POTENTIAL).
 CC FT ACT_SITE 292 292 PROBABLE.
 CC SQ SEQUENCE 1545 AA: 170627 MW: 35DD753988FD478 CRC64;

alignment_scores:
 Quality: 1587.00 Length: 1723
 Ratio: 1.710 Gaps: 47
 Percent Similarity: 53.860 Percent Identity: 26.291

alignment_block:

US-09-303-518d-653 x IGA3_HAEIN ..

Align seg 1/1 to: IGA3_HAEIN from: 1 to: 1545

```

64 CGCTTCTGCGCCGCTTACTTACCATATAGCCTGCTTGGCATTCTGCC 113
   ::::::::::: ::::::::::: ::::::::::: ::::::::::: ||
   5 LysPheLysLeuAsnPhelIleAlaLeuThrValAlaIleAlaLeuThrPr 21
114 CCAACCCGCGGCGGACACACTTATTGGGCATCACTACCAATCTCTC 163
   | ::::::::::: ::::::::::: ::::::::::: :::::::::::
21 oTyrThrGluAlaIleValAlaArgAspAspValAspTyrGlnIlePheA 38
164 GCGACTTGGCCGAAATAAGGCAAGTTGCGAGTCGGCGCAAGAATATT 213
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
38 rGAspPheAlaGluAsnLysGlyLysPheSerValGlyAlaIleThrAsnVal 54
214 GAGGTTTACACAAAAAGGAGGAGTGTGGCAAAATGATGACGAAAGC 263
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
55 GluValAlaArgAspLysAsnAsnArgProLeuGlyAsnValLeuProAsnG1 71
264 C...CGATGATGATTTTCTGTGTTATCG...CGTAAGGCGGTGGCGG 307

```

```

71 yIleProMetIleAspPheSerValValAspValAspLysArgIleAlaT 88
   : |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
308 CATTGGCGGCGGATCATATATATTGAGCGCTGGCAGAT...AACGGCGGC 354
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
88 hrLeuValAsnProGlnIleValAlaGlyValGlyLysHisLysSerAsnGly 104
335 TATAACATGTTGATTTGGT.....GCGGAGGCAACGAA 389
   ::::::::::: |||::|::|::|::|::|::|::|::|::|::|::|::|
105 ValSerGluLeuHisPheGlyAsnLeuAsnGlyAsnMetAsnAsnGlyAs 121
330 TCCCGATGACGCGC.....TTTTCTTAACCAATTTG 421
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
121 nAlaLysAlaHisArgAspValSerSerGluGluAsnArgTyrThrV 138
422 TGAAGAATATTAATAAACGAGGACTACGCG..... 456
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
138 alGluLysAsnGluTyrProThrLysLeuAsnGlyLysAlaValThrThr 154
457 .....CATCCTATGCGGCGGATATCATATGCGCGCTTGGCA 494
155 GluAspGlnAlaGlnLysArgArgLysAspTyrTyrMetProArgLeuAs 171
495 CAAATTGTGCAGATGCGAGAACCTGTTGAG...ATGACGATATATATG 541
171 pLysPheValThrGluValAlaProIleGluAlaSerThrAspSerSert 188
542 ATGGGTGGAATATACCGTGATTTAAATAATACCTGATCGTGTGGAATC 591
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
188 hrAlaGlyThrTyrAsnAsnLysAspLysTyrProTyrPheValArgLeu 204
592 GGACGAGCGACACATAT.....TG 611
205 GlySerGlyThrGlnPheIleTyrGluAsnGlyThrArgTyrGluLeuTr 221
612 CGGCTGTGATGACAGACACCAATTAACCGGAAAGTTCAATCAT... 657
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
221 pLeuGlyLysGluGlyGlnLysSerAsp...AlaGlyGlyTyrAsnLeuL 237
658 .....ATTGCAGGCGCATATTTGCGTCGCGGTGGCATACCTTTGCA 702
237 yLsLeuValGlyAsnAlaIleTyrThrTyrGlyIleAlaGlyThrProTyrGlu 253
703 CAAATGATCAGGTGGTGGCAGCATCACTTAGGTAGC..... 741
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
254 ValAsnHisGluAsnAspGlyLeuIleGlyPheGlyLysnSerAsnG1 270
742 .....GAAAAATTAAACATAGCCCATATGTTTATAC 775
270 uTyrIleAsnProLysGluIleLeuSerLysPro.....LeuT 284
776 CAACAGAGGCTCATTTGGGACAGTGGCTACCAAGTTATCTATGAT 825
284 hrAsnTyrAlaValLeuGlyAspSerGlySerProLeuPheValTyrAsp 300
826 GCCCAAAAGCAAAAGTGTAAATTAATGGGGTATTGCAACAGCAACCC 875
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
301 ArgGluLysGlyLysTrpLeuPheLeuGlySerTyrAspTyrTrpAlaG1 317
876 CTATATGGAAGAAAAGCAATGGCTTCAGCTAGTTGTAAGATGTTGTTCT 925
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
317 yTyr.....AsnLysLysSerTrpGlnG 325
926 ATGATGAATCTTGTGAGATACCATATCATATTTACGCAACACAT 975
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
325 IuTrpAsnIleTyrLysProGluPheAlaGluLysIleTyrGluGlnTyr 341
976 CAAATGGAATATCTTTTAAAGCAATAATATATGCGCAGCAAAAT 1025
342 .....SerAlaGlySerLe 346
1026 CGATGCCAAACATTAACACTATTCTACCTTATAGATTAAACACGAA 1075
   : ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

```

```

346 uileglyserlysthraspyrsertrpserasnglylysthrserT 363
1076 CCGTT.....CAATTGTTATGTTCTTTATCGAAGCA 1113
      |||:|||||:|||||:|||||:|||||:|||||:
363 hrllerhglyglulysserleuasnvalaspleualaaspglylys 379
1114 AGAAGCTGTTTATCATGTCGAGGTGGGTCAACAGTTATGACCCAG 1163
      |||:|||||:|||||:|||||:|||||:|||||:
380 AsplysPro..... 382
1164 ACTGAAATATGAGAAATATTCCTTATTGCAAGAAAGGTAAT 1213
      |||:|||||:|||||:|||||:|||||:|||||:
383 ...AsnhisglysserValthrPhe.....glulglyserglythrL 396
1214 TGATCTTACAGACATCAACCAAGCGGGCGGTGTTGTTATTTAG 1263
      |||:|||||:|||||:|||||:|||||:|||||:
396 eutThrLeuasnasnalleaspginglyalaglyleuphepheelu 412
1264 GTATATTTACGGTC...TCGCTAAAMCAACAAACGTTGCAAGCGC 1310
      |||:|||||:|||||:|||||:|||||:|||||:
413 GlyaspyrGlulValLysglyThrSeraspasnThrtrpLysglyAl 429
1311 GGGCGTTATTCAGTATGCGACGTACCGTTACTTGAAGTAACGGCG 1360
      |||:|||||:|||||:|||||:|||||:|||||:
429 aglyValserValalagluglylysthrValThrtrpLysValhisasp 446
1361 TGGCAAGACCGCCTGTCCAAATCGGCAAGGACGCTGCTGTTCAA 1410
      |||:|||||:|||||:|||||:|||||:|||||:
446 roglIntyrAspArgLeuAlalyslleglylysglyThrLeuileValislu 462
1411 GCCAAAGGGGAAACCAAGGCTCGGTACGCGTGGCGACGGTAAAGTCAT 1460
      |||:|||||:|||||:|||||:|||||:|||||:
463 glyThrGlyaspasnlysglyserLeuLysValGlyaspGlyThrValI 479
1461 CTTAATATGACAGGGGACGATCAAGCAAAACCAACGCTTTAGTGA 1510
      |||:|||||:|||||:|||||:|||||:|||||:
479 eueuLysingInthrasnlysergly...GlnhisAlapheLaserV 495
1511 TCGCGTTGTCAGCGGAGGGGAGCGTGAACCTGAATGCCGTAATATG 1560
      |||:|||||:|||||:|||||:|||||:|||||:
495 alglylleValiserGlyarSerThrLeuValleuasnaspaspLyslu 511
1561 TTCAAACCCGACAACTCTATTTGCGCTTTCGCGGCGGACGTTGGATT 1610
      |||:|||||:|||||:|||||:|||||:|||||:
512 ValasprProasnseriletyrPhegllyPheArglyArgLeuaspLe 528
1611 GAACGGGATTCGCTTTCGTTCCACCGATCAAAATACCGATGAAGGG 1660
      |||:|||||:|||||:|||||:|||||:|||||:
528 uasnilyasnserileutThrPheaspHisileArgasnilleaspgingly 545
1661 CGATGATTGTCAACCAATCAAGCAAAAGATCCACGTTACATTACA 1710
      |||:|||||:|||||:|||||:|||||:|||||:
545 laArgLeuValasnHisserThrserLysHisserThrValThrlethr 561
1711 GGCATTAAGATATTACTACAAACGGCAAT..... 1740
      |||:|||||:|||||:|||||:|||||:|||||:
562 GlyaspasnLeullethrAspProasnvalSeriletyrVally 578
1740 ..... 1740
578 sProleuGluaspaspasnProtyrAlalaleargInllelystyryglYT 595
1740 ..... 1740
595 yrglInleutyrrPheasnglugluasnArgThrtyrAlaleuLyslys 611
1740 ..... 1740
612 AspalaserlleArgSerGluPheProgluasnArgglyLysSerAsnas 628
1741 .....A 1741
628 nsertrpleutyrmetyrGlyThrGluLysAlaaspalagInlysaasnaliam 645
1742 ACAACACTTTGGATTACCAAAAAAGAAATTCCTACACAGCTGTTGGC 1791
      |||:|||||:|||||:|||||:|||||:|||||:
645 eLasnHisilleasnasngluArgmetasnglyPheasnnglytyrPhegl 661
1792 GAGAAAGATGCAACCAAAACGAACGGCGGCTCATATGTAATTACCAAC 1841
      |||:|||||:|||||:|||||:|||||:|||||:
662 Glugluglugly...LysasnasnnglyLysnleuAsnValThrPheLysgl 677
1842 GGAAGACGGGATCGCATTTACTGTTCCGGCGGCAAAATTTAAACG 1891
      |||:|||||:|||||:|||||:|||||:|||||:
677 LysSerGluGluAsnArgPheLeuLeuThrGlyglYthrAsnleuAsng 694
1892 GCAATATCACGCAACAAACGCAACGCAACTGTTTTCACGCGACAGCCACA 1941
      |||:|||||:|||||:|||||:|||||:|||||:
694 LysaspLeuasnValIngInglyThrLeuPheLeuSerGlyArgProthr 710
1942 CCGCAGCGCTACATCATTTAGAGACGGGTGGTCAAAATGGAAGT... 1989
      |||:|||||:|||||:|||||:|||||:|||||:
711 ProhisAlaArgaspIlealaglylleSerSerThrLysLysAspserHi 727
1990 ...ATCCCAACAGAGAAATGCTGTGGCAACACGATGTGATCGACCGCA 2035
727 sPheSerGluAsnAsnGluValalValalGluaspasprPleasnArgA 744
744 snPheLysAlatrhrasnilleasnValThrAsnasmalatrLeutyser 760
2086 ...CGCAATGTCGCAAAAGTGGAGGCGGATGGCATTTAAGCAATACGC 2132
761 GlyArgasnValalGlnserileThrSerasnilleThrAlaserAsnsmal 777
2133 CCAAGCAGTTTTCGTGTGTCACCGCATCAAAAGCCACATCTGTACAC 2182
      |||:|||||:|||||:|||||:|||||:|||||:
777 alysValHisilleglytyr....LysAlaglyAspThrValCysValA 792
2183 GTTCGACGTGGACGGGTCTGACAAAGTTGTACGAAAAAACATTACCGAC 2232
      |||:|||||:|||||:|||||:|||||:|||||:
792 rgserspyrThrGlytyrValThrCysThrThrAspLysLeuSer... 807
2233 GATTAAGTGATTCGTCATTTAGCAAGACCGCATGACAGGCGCAATGAC 2282
      |||:|||||:|||||:|||||:|||||:|||||:
808 AsplysAlaLeuasnserPheasnProthrAsnleuArgGlyAsnValas 824
2283 CCTTGGCGATCACGCTCATTTAAATCTCACAGGACTTGCACACTACAG 2332
      |||:|||||:|||||:|||||:|||||:|||||:
824 nleutThrGluSerAlaasnPheValleu...GlyLysAlaasnleuPheg 840
2333 GCAATCTTAAGTGCAGGCGGAGACACGACATATACGTTACGGCGCAACGCGC 2382
      |||:|||||:|||||:|||||:|||||:|||||:
840 LyrThrileGlnSerArgglyAsnserGlnValArgleuThrGluasn... 855
2383 ACCCAAAACGGCAACCTCAGCCTCGTGGCAATGCCCAAGCAACATTAA 2432
      |||:|||||:|||||:|||||:|||||:|||||:
855 ..... 855
2433 TCAAGCCACATTAAACGGCACACATCGCTTCGACAATGCTTCATTTA 2482
      |||:|||||:|||||:|||||:|||||:|||||:
855 ..... 855
2483 ATCTAAGCAACAACGCCGTACAAACGGCAGTGTGACGGCTTTCGACAAC 2532
      |||:|||||:|||||:|||||:|||||:|||||:
855 ..... 855
2533 GCTAAGCAACGTAAGCCATTCGACATCAAGCAATGTCCTCCAGC 2582
      |||:|||||:|||||:|||||:|||||:|||||:
855 ..... 855
2583 CGATTAAGCAGATTCCATTGTAACACGCGCTTACCGGAAAAATCA 2632
      |||:|||||:|||||:|||||:|||||:|||||:
855 ..... 855

```

```
2633 GCGGGCGCAAGATACGGCATTTACATTAAAGACGGAATGACGCTG 2682
856 .....SerHisTriphHisLeu 860
2683 CCGTGGGCGACGATTTAGCGAATTTAAACCTTGACAAAGCCACCATTAC 2732
861 ThrGlyAsnSerAspValHisGlnLeuAspLeuAlaAsnGlnHisIleH1 877
2733 ACTCAATTCGGCTATCGACAGATGGCGGCAAGCCGCAACCGGCATCG 2782
877 SLeuAsnSerAlaAspAsnSerAsn..... 885
2783 CGGCAGATGCGCGCGCGCGCTTGCAGCCGTTCCCTATTATCCGTACG 2832
886 .....AsnValThr 888
2833 CCGCCAACTCGCGCAGATCCCGTTTCAACAGCGTGAAGGTAACGGCA 2882
889 .....LysTyrAsnThrLeuThrValAsn...Se 897
2883 ATTGAACGGTCAAGGACATTCGGCTTATGCGAATCTTGCGGCTACG 2932
897 rLeuSerGlnAsnGlnSerPheTyrTyrLeuThrAspLeuSerAsnLysG 914
2933 GCAGCGCAATTTGAAGCTGGCGGAAGTTCGGAAGCACTTACACCTTG 2982
914 InGlnAspLysValValValThrLysSerAlaThrGlnLysAsnPheThrLeu 930
2983 GGTGCACAAATACCGGCAAGAACCCCTAGTCTCGAGCAATGACGCT 3032
931 GlnValAlaAspLysThrGlnLysPro...AsnHisAsnGlnLeuThrLe 946
3033 AGTGAAGGAAAGACACACACCGCTGTCGAAATCTTAATTTCAACG 3082
946 uPheAspAla....SerLysAlaGlnArgAspHisLeuAsnValSerL 961
3083 TGCAAAACGAACACGTCGATCGCGCGCAGCGCTTATCAGCTTATCCGC 3132
961 euValGlnAsnThrValAspLeuGlnAlaThrLysTyrLysLeuArgAsn 977
3133 AAGACGCGAGTTCGGCTGCAATATCCG..... 3162
978 ValAsnGlnArgTyrAspLeuTyrAsnProGlnValGlnLysArgAsnG1 994
3162 ..... 3162
994 nThrValAspThrThrAsnIleThrThrProAsnAsnIleGlnAlaAspV 1011
3162 ..... 3162
1011 alProSerValProSerAsnAsnGlnGlnIleAlaArgValAspGlnAla 1027
3162 ..... 3162
1028 ProValProProProAlaProAlaThrProSerGlnThrThrGlnThrVa 1044
3162 ..... 3162
1044 lAlaGlnAsnSerLysGlnGlnSerLysThrValGlnLysAsnGlnGlnAla 1061
3162 ..... 3162
1061 spAlaThrGlnThrThrAlaGlnAsnArgLysValAlaLysGlnValLys 1077
3163 .....GTCAAGAAACAAGAGCTTCGACAAACTCGGCAAGGGGGA.. 3204
1078 SerAsnValLysAlaAsnThrGlnThrAsnGlnValAlaGlnSerGlySe 1094
3205 .GAACAGAGGCGCTTGACGGCAAAACAGGCACTTGCGGCCCAAC 3253
1094 rGlnThrLysGlnThrGlnThrThrGlnThrLysGlnThrAlaThrValG 1111
3254 AACAGGCGGAAAGACACCGCGCAAGC..... 3282

1111 lLysGlnGlnLysAlaLysValGlnThrGlnLysThrGlnGlnValPro 1127
3283 ...CTTGACGCGCTGATTCGGCGCGGCGCAATGCCAACCGCAAGAGCAGA 3329
1128 LysValThrSerGlnValSerProLysGlnGlnSerGlnThrValG1 1144
3330 AAGTGTGCCGACCGCGCGGCAAGGCGGGAATTCGCGCATTA 3379
1144 nProGlnAlaGlnProAlaArg.....GluAsnAspProThrV 1157
3380 TGCAGCGCGAGAAAGAAAAACGGGTGACGCGGATAAGACACCGCC 3429
1157 alAsnIleLysGlnProGlnSerGlnThrAsnThrThrAlaAspThrGln 1173
3430 TTG...CGAAACAGCGCGAAGCG..... 3450
1174 GlnProAlaLysGlnThrSerSerAsnValGlnProValThrGlnSe 1190
3451 .....GAACCC 3457
1190 rThrThrValAsnThrGlnLysAsnSerValValGluAsnProGluAsnThrT 1207
3458 GCGCGGCTACACCGCCTTCCG..... 3480
1207 hrProAlaThrThrGln...ProThrValAsnSerGlnSerSerAsnLys 1222
3481 .....CGCGCGCGCGCGCGCGCGGATTTCCG...CAACGCA 3518
1223 ProLysAsnArgHisArgArgSerValArgSerValProHisAsnValG1 1239
3519 GCCCAACCGCAACCCCAACCGCAGCGC.....GACC 3550
1239 uProAlaThrThrSerSerAsnAspArgSerThrValAlaLeuCysAspL 1256
3551 TGATCACCGCTTATGCCAATAGCGGTTTGAGTAATTTCCGCGC..... 3594
1256 euThrSerThrAsnThrAsnAlaValLeuSerAspAlaThrGlnAlaLysAla 1272
3595 .....ACGCTCAACAGCGCTTTCCGCGGTACAGAGCAATTTGACCG 3635
1273 GlnPheValAlaLeuAsnValGlnLysAlaValSerGlnHisIleSerG1 1289
3636 CGTG...TTGCGGAGACCGCGCAACCGCTTGACAGCGGATCC 3682
1289 nLeuGlnMetAsnAsnGlnLysGlnTyrAsnValThrValSerAsnThrS 1306
3683 GGGACACCAACACATACCGTTGCAAGATTCGCGCGCTACCGCCAAACA 3732
1306 erMetAsnLysAsnTyrSerSerSerGlnTyrArgArgPheSerSerLys 1322
3733 ACCGACCTGCGCAATCGGTATGCAGAAAAACCTCGGACGCGG...CG 3779
1323 SerThrGlnThrGlnLeuGlnTyrAspGlnThrIleSerAsnAsnValG1 1339
3780 CGTGGCAGTCGTTGTTGCGACACCGGACCGGAACACCTTCGACGACG 3829
1339 nLeuGlnGlnValAlaPheThrTyrValArgAsnSerSerAsnAsnPheAspLysA 1356
3830 GCATTCGCAACTCGGACGCGCTTGCCACAGGTGCTTTTCGGGCAATAC 3879
1356 lThrSerLysAsnThrLeuAlaGlnValAsnPheLysThrLysTyrTyr 1372
3880 GGCATTCGACGTTGACATCGGCATCAGCGCGCGCGGCTTTTAAGTAG 3929
1373 AlaAspAsnHisTyrTyrLeuGlnIleAspLeuGlnTyrGlnLysPheG1 1389
3930 CCGCAGCCTTTCAGACGCGCATGACGAGGCAAAATCGCGCGCGCGTGC 3979
1389 nSerLysLeuGlnThrAsnHisAsnAlaLysPheAlaArgHisThrAlaG 1406
3980 ATTACGCGATTGACGCAAGATACCGCGCAGTTTCGGCGATTTCGCGATC 4029
```

```

1406 InPhegIleuThraIaGlySalaPheAsnLeuGlyAsnPhegIyle 1422
4030 GAACCGCATGGCGCAACGGCGTATTGCGCAAAAAGGATTCAGG 4079
1423 ThrProIleValIaIArgTyrSerTyrLeuSerAsnAlaAspPheAl 1439
4080 ATACGAACAGTCATATTCGCCACCGCGGCGCTTCATTCACGCTAC 4129
1439 aleuAspIlnAlaIArgIleLysValAsnProIleSerValIysThrAlaP 1456
4130 GCGCGGCGATTAAGCAGCATATTCATTCACCAACCGCGGCACACATTCC 4179
1456 heaIaGlnValaIAspLeuSerTyrThrIAsnValIAsnGlyGlu...PheSer 1471
4180 ATACAGCGCTTATTTGCGCTGCTCATACCGATCGCGCTCGGCAAGT 4229
1472 ValThrProIleLeuSerAlaIArgTyr...AspAlaAsnGlnGlySerGI 1487
4230 CCGAAGCGCGCTCATACCGCGCGTATTCGCGCAGATTTCGCAAAACCC 4279
1487 yLysIleAsnValaIAsnGlyTyrAspPheAlaIAsnValaIAsnGlnG 1504
4280 GAGAGCGGGAATGGGCGTAAACGCCGAATCAAAAGTTTCACGCTGTC 4329
1504 InGlnIyrAsnAlaGlyLeuLysLeuLysTyrHisAsnValIAsnLeuSer 1520
4330 CTCACGCTGCGCGCGCCAGCGCGCATTCGCAATGGAGCGCAGACAGCGC 4379
1521 LeuIleGlyGlyLeuThrLysAlaLysGlnIaGlyGlnLysThrAl 1537
4380 GGGCATCAATTAGCTAC 4398
1537 aGluLeuLysLeuSerPhe 1543
seq_name: SwissProt_40:IGA2_HAEIN
seq_documentation_block:
ID IGA2_HAEIN STANDARD; PRT; 1702 AA.
AC P45384;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Immunoglobulin A1 protease precursor (EC 3.4.21.72) (ICAI protease).
GN IGA.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN 11
RP SEQUENCE FROM N.A.
RC STRAIN=HK715 / SEROTYPE B.
RX MEDLINE=92234949; PubMed=1373717;
RA Poulsen K., Reinholdt J., Kilian M.;
RT "A comparative genetic study of serologically distinct Haemophilus
RT influenzae type 1 immunoglobulin A1 proteases."
RL J. Bacteriol. 174:2913-2921(1992).
CC -1- FUNCTION: VIRULENCE FACTOR; CLEAVES HOST IMMUNOGLOBULIN A
CC PRODUCING INFECT FC AND FAB FRAGMENTS.
CC -1- CATALYTIC ACTIVITY: Cleavage of immunoglobulin a molecules at
CC certain Pro-I-Xaa bonds in the hinge region. No small molecule
CC substrates are known.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- DOMAIN: THE SIGNAL PEPTIDE GUIDES THE PRECURSOR TO THE PERIPLASMIC
CC SPACE, AND THE CARBOXY-TERMINAL HELPER DOMAIN ASSOCIATES WITH THE
CC OUTER MEMBRANE TO FORM A PORE FOR EXCRETION OF THE PROTEASE
CC DOMAIN. THE HELPER DOMAIN IS THEN RELEASED BY AUTOPROTEOLYSIS (BY
CC SIMILARITY).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S6 (SERINE PROTEASE).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

```

```

CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M87489; AAA24966.1; -.
DR MEROPS: S06.001; -.
DR InterPro: IPR000710; IGA_S6.
DR Pfam: PF02395; IGA1; 1.
DR PRINTS: PR00921; IGASERPTASE.
KW Hydrolase; Serine protease; Transmembrane; zymogen; Repeat; Signal.
FT SIGNAL 1 25
FT CHAIN 26 1014 POTENTIAL.
FT PROPEP 1015 1702 IMMUNOGLOBULIN A1 PROTEASE.
FT ACT_SITE 288 288 HELPER PEPTIDE (POTENTIAL).
FT DOMAIN 1109 1124 PROBABLE.
FT 2 x 8 AA TANDEM REPEATS OF A-K-V-E-K-E-E-
FT REPEAT 1109 1116 K.
FT REPEAT 1117 1124 1.
FT REPEAT 1117 1124 2.
SQ SEQUENCE 1702 AA; 186539 MW; 860F70D2667807A6 CRC64;

```

```

alignment_scores:
Quality: 1583.50 Length: 1867
Ratio: 1.679 Gaps: 48
Percent Similarity: 50.509 Percent Identity: 25.174

```

alignment_block:

US-09-303-518D-653 x IGA2_HAEIN ..

Align seg 1/1 to: IGA2_HAEIN from: 1 to: 1702

```

64 CGCTTCGCGCGCGCTTACTTACGCAATATGCTGCTGGCATTCGCC 113
5 LysPheLysLeuAsnPhelIleAlaLeuThrValAlaIAsnValaLeuThrPr 21
114 CCAAGCGCGGCGGACACACTATTTCGCGATCAGCTCCATCATATC 163
21 oTyrThrGlnAlaIAlaLeuValaIArgAspAspValaIAspTyrGlnIlePheA 38
164 GCGACTTTCGCGCAAAATTAAGCGAAGTTTGCAGTCGCGCGAAGATAT 213
38 rGAspPheAlaIAsnValaIAsnLysGlyAArgPheSerValaIAsnValaI 54
214 GAGGTTTACACAAACAAAGGAGGAGTTCGCGCAATTCATGACGAAC 263
55 GluValaIArgAspLysAsnAsnHisSerLeuGlyAsnValaLeuProAsnI 71
264 C...CGATGATGATTTTCTGCTGATGCG...CCTAACGGCGTGGCGG 307
71 yLleProMetIleAspPheSerValaIAspValaIAspLysArgIleAlaI 88
308 CATTGGCGGCGATCATATATTTGAGCGTGGACAT...AACGGCGG 354
88 hrLeuIleAsnProGlnTyrValaIValaIValaIYshIAsnValaIsera 104
355 TATAACAATGTGATTTTGT.....GCGGAGGAGGACAA 389
105 ValSerGluLeuAsnIspheGlyAsnLeuAsnGlyAsnMetAsnIAsnLys 121
390 TCCGCATCAGCAGCGC.....TTTCTTACCAATG 421
121 naPlySerHisArgAspValaIserSerGluGlnaAsnArgTyrPheSerV 138
422 TGAAGAAGAAATTAATTAAGCAGGAGACTAACGCG..... 456
138 aGluLysAsnGlnTyrProThrLysLeuAsnLysLysAlaValThrThr 154
457 .....CATCCTATGCGCGCATATATCATATGCGGCTTGA 494
155 GluAspGlnThrGlnLysArgArgGlnAspTyrTyrMetProArgLeuAs 171
495 CAATTTGTCAGATGACAGACAGCTGTGAG...ATGACAGATTATATG 541
|||||

```



```
171 pLysPheValThrGluValAlaProIleGluAlaSerThrAlaSerSer 188
542 ATGGGTGGAATACGGCTGATTAAATTAATACCGATCGCTGTTCGAATC 591
188 sPrAlaGlyThrTyrAsnAspGlnAsnLysTyrProAlaPheValArgLeu 204
592 GGAAGCAGCAGACAAATATTGGCGGTCTGATGAAGAGAA..... 630
205 GlySerCysSerGlnPheIleTyrLysGlyAspAsnTyrSerLeuI 221
631 .CCCAATACCGCGAAAGTTCAATAT.....ATTGCAAGC 667
221 eLeuAsnAsnHisGluValGlyLysAsnLeuLysLeuValGlyAsp 238
668 CATATTCTGGCTGCTGGTGGCAATACCTTTGCACAAATGATGATGAGT 717
238 lArgThrTyrGlyIleAlaGlyThrProTyrLysValAsnHisGluAsn 254
718 GGTGGCAGACGTCACTTAGGTAGCAGAAATAATTAACATAGC...CCATA 764
255 AsnGlyLeuIleGlyPheGlyAsnSerLysGluGlnHisSerAspPro 271
765 TGCTTTTACCAACAGCA.....GGCTATTGGCGACA 799
271 sGlyIleLeuSerGlnAspProLeuThrAsnTyrAlaValIleGlyAsp 288
800 GTGGCTCACCAATGTTTATCTATGATGCCCAAGCAAAAGTGTTAAAT 849
288 eArgLysProLeuPheValTyrAspArgGlnLysGlyLysTyrLeuPhe 304
850 AATGGGTATTGCMAACAGCAACCCCTATATAGCAAAAGCAATGGCTT 899
305 LeuGlySerTyrAspPheTyrAlaGlyTyrAsnLysLysSerTyrGln 321
900 CCAGCTAGTTCGTAAGATTGGTCTATGATGAATCTTGTCTGAGATTA 949
321 uTrrAsnIleTyrLysProGluPheAlaLysThrValLeuAspLysAsp 338
950 CCCATTCACTATTCTAGCAACCAATCAAAATGCAAAATCTTTTAAAC 999
338 hr..... 338
1000 GACATTAATATGGCGCAGGAAATAATCGATGCCAAACATTAACATA 1049
339 .....AlaGlySerLeuThrGlySerAsnThrGlnTyrAs 350
1050 TCTACCTTATAGATTAAACACGAAACGCT.....CAATGT 1087
350 uTrrAsnProThrGlyLysThrSerValIleSerAsnLysSerGluSer 367
1088 TTAATGTTCTTTATCCGAGACAGACAGAAACCTGTTATCATGCTGCA 1137
367 euAsnValAspLeuPheAspSerSerGlnAsp..... 377
1138 GGTGGGTGCACAGTTATCGACCCAGACTGAATATATGAGAAATATTC 1187
378 .....ThrAspSerLysLysAsnHisGlyLysSerValThr 390
1188 CTTTATGACAAAGGAAAGTGAATGATTAATACGCAACATCAAC 1237
390 rLeu.....ArgLysSerGlyThrLeuThrLeuAsnAsnHisLeasp 405
1238 AAGCGCGCGGCGTTGATTATTGAGGTAATTTTACGCTC...TCGCGT 1284
405 lArgIleGlyGlyLeuPhePheGlyLysAspTyrGluValLysGlyThr 421
1285 AAAAACAACGAAACGTGGCAAGCGCGGCTTCATATCAGTGAAGGAG 1334
422 SerAspSerThrThrLysGlyAlaGlyLysValAlaAspGlyLys 438
1335 TACGCTTACTTGGAAAGTAAAGCGCGTGGCAACGACCGCTGTCAAAA 1384
438 sThrValThrTrrLysValHisAsnProLysSerAspArgLeuAlaLys 455
1385 TCGCAAAAGCAGCGCTCTGTTCAGCCAAAGGGAAGCAACAGCGCTG 1434
455 lGlyLysGlyThrLeuIleValGlnGlyLysGlyLysGlnLysGlySer 471
1435 GTACAGCGGCGGCGGTAAGTCACTTAGATCGACGCGGAGCATCA 1484
472 LeuLysValGlyAspGlyThrValIleLeuLysGlnGlnAlaAspAla 488
1485 AGGCAAAACAGACCTTATGAAATCGCTTGCTGAGCGGAGGAGGA 1534
488 AsnLysValLysAlaPheSerGlnValGlyIleAlaSerLysArgSer 505
1535 CGGTGCAACTGATCCGATATACGTTCAACCCGACAAACCTATTTTC 1584
505 hrValValLeuAsnAspAspLysGlnValAspProAsnSerIleTyrPhe 521
1585 GCGTTTCGCGGCGAGCTTGGATTGAAAGCGGCAATCGCTTGCTGCA 1634
522 GlyPheAlaGlyGlyArgLeuAspAlaAsnGlnLysAsnLeuThrPheG 538
1635 CCGCATTCAAAATACCGATGAAGGCGCGATGATGTCACCAACATCAAG 1684
538 uHisIleArgAsnIleAspArgLysAlaArgLeuValAsnHisAsnThr 555
1685 ACAAGAATCCACCGTTACCTTACAGGCAATTAAGATATTACT..... 1728
555 eLysThrSerThrValThrIleThrGlyLysSerLeuIleThrAspPro 571
1728 ..... 1728
572 AsnThrIleThrProTyrAsnIleAspAlaProAspGlnAspAsnPro 588
1728 ..... 1728
588 rAlaPheArgArgIleLysAspGlyGlnLeuTyrLeuAsnLeuGluA 605
1728 ..... 1728
605 sNTrThrTyrTyrAlaLeuArgLysGlyAlaSerThrArgSerGluLeu 621
1729 .....ACMACCGGCAATTAAC..... 1743
622 ProLysAsnSerGlyLysSerAsnGlnAsnTrrLeuTyrMetGlyLys 638
1744 .....AACMACTTGATGCAAAAAAG 1765
638 rSerAspGlnAlaLysArgAsnValMetAsnHisIleAsnAsnGlnArg 655
1766 AATTGCTACAAACGTTGGTGGCGAAGATGCAACCAACCAACGAAAC 1815
655 eTanGlnPheAsnGlyTyrPheGlyGlnGlnGlyLysAsnAsn 670
1816 GGGCGGCTCAATCGAATTAACCAACCGAAGAACCGGATGCGCACTTACT 1865
671 GlyAsnLeuAsnValThrPheLysGlyLysSerGlnGlnAsnArgPhe 687
1866 GCTTCGCGGGAACAATTAACGGCAATATACGCAACCAACGCGCA 1915
687 uLeuThrGlyGlyThrAsnLeuAsnGlyAspLeuLysValGlyLysGly 704
1916 AACTGTTTTCAGCGCAGACGACACCGCAGCTACATCAATTAAGGA 1965
704 hrLeuPheLeuSerGlyArgProThrProHisAlaArgAspIleAlaGly 720
1966 AGCGGTGCTCAAAATGCAAGT.....ATCCACAAAGGAGAAATCGT 2009
721 lIleSerSerThrLysAspGlnHisPheAlaGlnAsnGlnValVal 737
2010 GTGGACAAACGATTGGATGACCGCAATTAAGGGAAGCAAACTTCGATA 2059
737 lValGlnAspAspTrrIleAsnArgAsnPheLysAlaTrrAsnIleAsn 754
```



```

1221 LuArgLysThrAlaLysValGluThrGluThrGlnGluProPro 1237
3340 GAACCGCGCGCGGCGAGC..... 3360
1238 GlnValAlaSerGlnAlaSerProLysGlnGluGlnSerGluThrValG1 1254
3361 .....GGGAAATGCC...GGCATTTACGAGCGG 3388
1254 nProGlnAlaValLeuGlnSerGlnValProThrValAsnAlaG 1271
3389 AGGAAGAGAAAAACGGGTGCGAGCGGATTAAGAC..... 3423
1271 LuGlnValGlnAlaGlnLeuGlnThrGlnThrSerAlaThrValSerThr 1287
3424 .....ACGGCCTT 3431
1288 LysGlnProAlaProGlnAsnSerIleAsnThrGlySerAlaThrAla11 1304
3432 GCGGAAACGCGCGAAGCGGAAACCGCGCGCTAC..... 3468
1304 eThrGluThrAlaGluLysSerAspLysProGlnThrGluThrAlaAla 1321
3468 ..... 3468
1321 eThrGluAspAlaSerGlnHisLysAlaAsnThrValAlaAspAsnSer 1337
3469 .....ACGCGTTCCCGCGCGCGCGCGCGCGCGCG 3500
1338 ValAlaAsnAsnSerGlnSerGluProLysSerAlaArg...ArgArg 1353
3501 GGATTTGCGCGCAACGCGAG..... 3519
1353 gSerIleSerGlnProGlnGluThrSerAlaGluGluThrThrAlaAla 1370
3520 .....CCCCAACCGCA 3531
1370 eThrAspGluThrThrIleAlaAspAsnSerLysArgSerLysProAsn 1386
3532 CCCCACCGCAG..... 3543
1387 ArgArgSerArgArgSerValArgSerGluProThrValThrAsnGlySe 1403
3544 .....CGGACCTGATCAGCGCTATGCCAATA 3571
1403 AspArgSerThrValAlaLeuArgAspLeuThrSerThrAsnThrAsnA 1420
3572 GCGGTTTGAGTGAATTTCCGCC.....ACGCTCAACAGC 3606
1420 LaValIleSerAspAlaMetAlaLysAlaGlnPheValAlaLeuAsnVal 1436
3607 GTTTTCGCCGTACAGACGAATGGACGCGCTG...TTTGGCGAAGACG 3653
1437 GlyLysAlaValSerGlnHisIleSerGlnLeuGluMetAsnGlnG1 1453
3654 CCGCAACGCGCTTGGACAGCGGATCCGGGACCAACACATACGCTT 3703
1453 yGlnIlyAsnValIlyPValSerAsnThrSerMetAsnGlnAsnIlySerS 1470
3704 CGCAAGATTTCCGCGCTACCGCAACAAACGACCTCGCAAAATCGGT 3753
1470 eSerGlnIlyArgArgPheSerSerLysSerThrGlnThrGlnLeuGly 1486
3754 ATGCGAAGAAACCTGCGACGCGG...CGCGTGGCATCGTTTTCGA 3800
1487 TrpArgGlnThrIleSerAsnAsnValGlnLeuGlyValPheThrTy 1503
3801 CAACCGACCGGAAACACCTTCGACGAGGATGGGCAACGCGACCGC 3850
1503 rValArgAsnSerAsnAsnPheAspLysAlaSerSerLysAsnThrLeuA 1520
3851 TTGCCACGCGTGGCGTTTGGGGAATACGACATCGCGAGGTTCGACAT 3900
1520 laGlnValAsnPheTySerLysTyGlyAlaAspAsnHisTrpLysLeu 1536

```

```

3901 GGCATCACCGCGCGCGGCTTTAGTAGCGGCAACGCTTTCAGCGCAT 3950
1537 GlyIleAspLeuGlyTyArgLysPheGlnSerAsnLeuLysThrAsnH1 1553
3951 CAGAGCGAAATCCGCGCGCGCTGCATTCAGGATTCAGCAAGAT 4000
1553 sAsnAlaLysPheAlaArgHisThrAlaGlnPheGlyLeuThrAlaGlyL 1570
4001 ACCGCGCAGTTTCGCGGATTCGCAATCGAACCGCAATCGCGCAACG 4050
1570 ySalPheAsnLeuGlyAsnPheGlyIleThrProIleValGlyValArg 1586
4051 CGCATTTTCGTCGCAAAAGCGGATTCAGGATTCAGCAAAACGTCATTC 4100
1587 TySerTyLeuSerAsnAlaAsnPheAlaLeuAlaLysAspArgIleLy 1603
4101 CACCCGCGCGCTTCATTCACCGCTACCGCGCGGCGATTAAAGCGCAT 4150
1603 sValAsnProIleSerValIlyThrAlaPheAlaGlnValAspLeuSerT 1620
4151 ATTCATTCAACCGCGCGCACACATTCATTCACCGCTTATTGAGCCTG 4200
1620 yThrTyHisLeuGlyGlu...PheSerValThrProIleLeuSerAla 1635
4201 TCCAT...ACCGATGCCGCTTCGCGCAAGTCGCAACGCGCGCATAC 4247
1636 ArgTyAspThrAsnGlnGlySerGlyLysIle....AsnValAsnG1 1650
4248 CGCGGTATTCGCGCAGATTCGCGCAAAACCGCGAGTCGGAATGGGCG 4297
1650 nTyAspPheAlaTyAsnValGluAsnGlnGlnIlyTyAsnAlaGlyL 1667
4298 TAAACCGCGAATCAAGGTTTACGCTGTCCTCCACGCTGCGCGCGC 4347
1667 euLysLeuLysTyHisAsnValLysLeuSerLeuIleGlyGlyLeuThr 1683
4348 AAGGCGCGCAATTCGAACCGCAGACAGCGCGCGCTCAATTAAGCTA 4397
1684 LysAlaLysGlnAlaGluLysGlnLysThrAlaGluLeuLysLeuSerPh 1700
4398 C 4398
1700 e 1700

seq_name: SwissProt_40:IGA_NEIGO
seq_documentation_block:
ID IGA_NEIGO STANDARD; PRT; 1532 AA.
AC P09790;
DT 01-MAR-1989 (rel. 10, created)
DT 01-MAR-1989 (rel. 10, last sequence update)
DT 16-OCT-2001 (rel. 40, last annotation update)
DE IGA-specific serine endopeptidase precursor (EC 3.4.21.72) (IGA
DE protease).
DE GN
OS Neisseria gonorrhoeae.
OX Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
RN NCBI_TaxID=485;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN=MS11;
RX MEDLINE=87115823; PubMed=3027577;
RA Pohlner J., Halter R., Beyreuther K., Meyer T.F.;
RT "Gene structure and extracellular secretion of Neisseria gonorrhoeae
RT IGA protease."
RL Nature 325:458-462(1987).
RN [2]
RP ACTIVE SITE.
RX MEDLINE=90154052; PubMed=2105953;
RA Bachovchin W.W., Plaut A.G., Flenk G.R., Lynch M., Kettner C.A.;
RT "Inhibition of IGA1 proteinases from Neisseria gonorrhoeae and
RT Hemophilus influenzae by peptide boronic acids."

```

RL J. Biol. Chem. 265:3738-3743(1990).
 CC -1- FUNCTION: THIS PROTEASE IS SPECIFIC FOR IMMUGLOBULIN A.
 CC -1- CATALYTIC ACTIVITY: Cleavage of immunoglobulin a molecules at
 CC certain Pro-I-xaa bonds in the hinge region. No small molecule
 CC substrates are known.
 CC -1- SUBCELLULAR LOCATION: Secreted.
 CC -1- DOMAIN: THE SIGNAL PEPTIDE GUIDE THE PRECURSOR TO THE PERIPLASMIC
 CC SPACE, AND THE CARBOXY-TERMINAL HELPER DOMAIN ASSOCIATES WITH THE
 CC OUTER MEMBRANE TO FORM A PORE FOR EXCRETION OF THE PROTEASE.
 CC DOMAIN: THE HELPER DOMAIN IS THEN RELEASED BY AUTOPROTEOLYSIS.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S6 (SERINE PROTEASES).
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC at the European Bioinformatics Institute. There are no restrictions on
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC

DR	EMBL; X04835; CAA25638.1; -.	
DR	PIR; A26039; A26039.	
DR	MEROPS; S06.001; -.	
DR	InterPro; IPR000710; IGA_S6.	
DR	Pfam; PF02395; IGA1; 1	
DR	PRINTS; PR00921; IGASERPTASE.	
KW	Hydrolase; Serine protease; Zymogen; Autocatalytic cleavage;	
KW	Transmembrane; Signal.	
FT	SIGNAL	1 27
FT	CHAIN	28 986
FT	PROPER	987 1532
FT	ACT_SITE	278 278
FT	SITE	986 987
FT	SITE	1018 1019
FT	SITE	1121 1122
SO	SEQUENCE	1532 AA; 168976 MW; 68FE4112BD22F40D CRC64;

alignment_scores:

Quality:	1565.50	Length:	1738
Ratio:	1.757	Gaps:	46
Percent Similarity:	51.266	Percent Identity:	27.100

US-09-303-518D-653 X IGA_NEIGO

Align seg 1/1 to: IGA_NEIGO from: 1 to: 1532

49 AAAACCGGGCGATCGCGTTCCGGCGCGCTTACTTACGACATATCCGTC 98
2 Lysalalysarphenyllysilleasnallalserleusertillephenual 18
99 GTTCGCATTTCGCCCAAGCCCGGGGACACACTTATTTCGGCATCA 148
18 atyrallaleuthrprtyrserclualalaleualalargapspsayala 35
149 ACTACCAATATGATCGCGACTTTGGCGAAATAAAGCAAGTTTGCAGTC 198
35 splyrglnllepheargpspheaialguasnulsglylysphephenal 51
199 GGGGCCAAGATATTGAGTTTACACAAAAAGGGAGTTGGTCGGCAA 248
52 GlyalalthraspseuserValLysasnulysarlgelgnasnilleglyas 68
249 ATCGATGACGAAGACCCCGCATGATTGATTTCTGTGTATCGCGTAC. 297
68 nalaleuserasrivalPrometillepspheneserValalaspvalasnL 85
298 .GGCGGCGCGCATTTGGCGGGCGATCATATATGTAGCGTGGACAT 345
85 ysarlgillealthrValaslaproglntYalalavalserVallyushis 101
446AACGGCGGTAAACAATGTTGA 368

[illegible]

```

344 ...GlyThrAsnSerHisIleGlySerThrAlaValArgLeuAlaAsnAs 359
1168 .....AATATGGAGAAATATTTCTTATTGACAAAGAA 1204
359 nGluGlyAspAlaAsnAsnGlyGlnAsnValThrPheGluAsp.....A 374
1205 AAGGTGAATTGATTAATTACGACGACATCAACCAAGCGCGCGGTTTG 1254
374 snGlyThrLeuValIleuAsnGlnAsnIleAsnGlnGlyAlaGlyLeu 390
1255 TATTTGAGGTAATTTTACGGCTCGCCCTAAACACAGAA...ACGTG 1301
391 PhePheSerGlyAspPyrThrValIleGlyAlaAsnAsnAspIleThrTr 407
1302 GCAAGCGCGCGGCTTCATATCATGTAGTACGATACCGTTACTTGGAAAG 1351
407 PheGlyAlaGlyIleAspValAlaAspGlyLysValValTrrGln 424
1352 TAAACGGCGGTGGCAACGACCGCGCTGCCAAATCGGCAAGGACGCTG 1401
424 AllysAsnProAsnGlyAspArgLeuAlaIleGlyLysGlyThrLeu 440
1402 CTGGTTCAAGCCAAAGGGAACCAAGCGTGGTACGCGTGGCGACGG 1451
441 GluIleAsnGlyThrGlyValAsnGlnGlyGlnLeuLysValGlyAsp 457
1452 TAAAGTCATCTTATGATCAGCAGCGGACGATCAAGGCAAAACAGACCT 1501
457 yThrValIleLeuAsnGlnIleAsnAspAlaAspLysLysValGlnAla 474
1502 TTGAGTGAATCGGCTTGGTACGCGGACGCGGACGCTGCACTGAATGCC 1551
474 hSerGlnValGlyIleValSerGlyArgGlyThrLeuValLeuAsnSer 490
1552 GATATACAGTTCAACCCGACAACTCTATTTCGCGCTTCGGCGGAGAG 1601
491 SerAsnGlnIleAsnProAspAsnLeuYrPheGlyPheArgGlyValYr 507
1602 TTGGGATTGGAACGGGCAATTCGCTTCGTTCCACCGCATTAATAATCCG 1651
507 GLeuAspAlaAsnGlyAsnAspLeuThrPheGlnHisIleArgAsnVal 524
1652 ATGAAGGGCGATGATGTCACACCAATCAACAGACAAATCCACCGT 1701
524 speIuGlyAlaArgIleValAsnHisAsnThrAspHisAlaSerThrIle 540
1702 ACCATTACGCAATTAAGATATTACTACACCGCG..... 1737
541 ThrLeuThrGlyLysSerLeuIleThrAsnProAsnSerLeuSerValHi 557
1737 ..... 1737
557 sSerIleGlnAsnAspTyrAspGluAspAspTyrSerTyrTyrArgP 574
1737 ..... 1737
574 roArgArgProIleProGlnGlyLysAspLeuTyrTyrLysAsnTyrArg 590
1737 ..... 1737
591 TyrTyrAlaLeuLysSerGlyLysArgLeuAsnAlaPrometProGluAs 607
1738 .....AATAAC..... 1743
607 nGlyValAlaGluAsnAsnAspTrrIlePheMetGlyTyrThrGlnGlu 624
1744 .....AACACTGTGATAGCAAAAGAAATTTGCC 1773
624 LuAlaArgLysAsnAlaMetAsnHisLysAsnAsnArgArgIleGlyAsp 640
1774 TACAACGGTGTGGCGAGAAAGATGCAACCAAGCAAGGCGCGCT 1823
641 PheGlyGlyPhePheAspGluGluAsnGlyLysGlyHisAsnGlyAlaLe 657
1824 CAATCTGAATTTACCAACCGGAAGACGGATCGCATTTACTGCTTCCG 1873
657 uAsnLeuAsnPheAsnGlnGlyLysSerAlaGlnLysArgPheLeuLeuThr 674
1874 GCGGAACAAATTTAAAGCGCATATACGCAACCAAGCGCAATCTGTT 1923
674 yGlyAlaAsnLeuAsnGlnGlyLysIleSerValThrGlnGlyAsnValLeu 690
1924 TTCACGGCGAGACGACACCGCACGCTTACATCATTTAGGAAGCGGCT 1973
691 LeuSerGlyArgProThrProHisAlaArgAspPheValAsnLysSerSe 707
1974 CTCAAAATGGAAGT.....ATCCCAAGAGAGAAATCTGTGGACA 2017
707 rAlaArgLysAspAlaHisPheSerLysAsnAsnIleValIlePheGlu 724
2018 ACGATTGGATGACCGCACATTTAAAGCGGAAACTCTCATATT...CAG 2064
724 sPAspTrrIleAsnArgThrPheLysAlaAlaGluIleAlaValAsnGln 740
2065 GCGGACAAAGCGGTGTTCCCGCAATGTGCCAAAGTGGAAGCGATTTG 2114
741 SerAlaSerPheSerSerGlyArgAsnValSerAspIleThrAlaAsnI 757
2115 GCATTTAAGCAATCAGCCCAAGCAGTTTGGGTGTCCACCGCATCAA 2164
757 eThrAlaThrAspAsnAlaLysValAsnLeuGlyTr.....LysAsn 772
2165 GCCACACAATCTGTACAGCTTCGAGCTGAGCGGCTGACAAAGTTGAC 2214
772 yAspGlyValLysValAlaArgSerAspTyrThrGlyTyrValThrCysAsn 788
2215 GAAAAACCATTTACCGACGATAAGTGATTCCTTCATTGACAGAACCGA 2264
789 ThrGlyAsnLeuSer...AspLysAlaLeuAsnSerPheAspAlaThr 804
2265 CATCAGAGCAATGTACAGCTTCGACGATCAGCTCATATTAAATCTCAG 2314
804 gIleAsnLysAsnValAsnLeuAsnGlnAsnAlaAlaLeuValLeu...G 820
2315 GACTTGCCACACTCAACGCGCAATCTTAGTGACGAGCGGAGACACGACTAT 2364
820 yLysAlaAlaLeuThrPheLysIleGlnGly..... 830
2365 ACGTTTACGCCCAACGCCACCAAAAGGCGACCTCAGCTCGTGGCAA 2414
831 ..... 831
2415 TGCCCAAGCAATTTAATCAAGCACATTTAAAGGCAACATCGGCTT 2464
833 nSerArgValSerLeuAsnGln..... 840
2465 CGGACATGCTTCATTTAATCTAAGCAACACGCGGTACAAAGCGCAGT 2514
840 ..... 840
2515 CTGACGCTTCCGACACGCTAAGCAACGTAAGCATTCGCGACTCAA 2564
840 ..... 840
2565 CGGCAATGTCTCCTAGCCGTAAGGCGATTTCCATTTGAAAACAGCG 2614
840 ..... 840
2615 GCTTTACCGGAAATATCAGCGGCGGCAAGATACGGCATTACACTTAAA 2664
840 ..... 840
2665 GACAGCAATGGAACGCTGCGCGTGGGCGACGATTTAGCAATTTAACT 2714
841 HisSerIleThrPheLeuThrGlyAspSerGlnValHisAsnLeuSerIle 857

```

```

2715 TGACAGCCGACGACATTCACATTCGCGCTATGACAGATGCGGACG 2764
      |||||
857 uAlaAspSerHisIleHisLeuAsnAsnAla.....SerAspAlaGlns 872
2765 GCGGCGAAACCGGACATGCGGACAGATGCGCGCGCGCTTCGCGCGT 2814
      |||||
872 erAla..... 873
2815 TCSCCTATTATCGGTTACGCGGCAACTTCGCGAGATCCGTTTCAACAC 2864
      |||||
874 .....AsnLysTyrHisIle 878
2865 GCTGACGGTAACGGCAATTTGAACGGTCAGGAAACATTCGCGCTTATGT 2914
      |||||
878 rIleLysIleAsn...HisLeuSerGlyAsnGlnHisIleHisTyrLeuT 894
2915 CGGAACCTTCGCGCTACCGGACGCGCAATTTGAAGCTTCGCGAAAGTTC 2964
      |||||
894 hIAspLeuAlaLysAsnLeuGlyAspLysValLeuValLysGluSerAla 910
2965 GAAGGCACTTACACCTTCGCTGACAAATACCGGCAACGACCGGTAAG 3014
      |||||
911 SerGlnHisTyrGlnIleuHisValGlnAsnLysThrGlyGluProAsnG 927
3015 TCCTGACGAATTTGACGCTAGTGGAAAGCAACACACACCGCTGTCC. 3063
      |||||
927 n.....GluGlyLeuAspLeuPheAspAlaSerS 937
3064 .....GAAATCTTAATTTTACCGCTGCAAAACGAAACAGCTC 3099
      |||||
937 erValGlnAspArgSerArgLeuPheValSerLeuAlaAsnHisTyrVal 953
3100 GATGCGCGGCGATGCGCTTATCAGCTTATCCGCAAGACGCGGAGTTCCG 3149
      |||||
954 AspLeuGlyAlaLeuArgTyrThrIleLysThrGluAsnGlyIleThrAr 970
3150 CCGTCATAT.....CCGCTCAAA..... 3168
      |||||
970 gLeuTyrHisnProTyrAlaGlnAsnGlyArgProValLysProAlaProS 987
3169 .....GAACACGAGCTTTCGCAACAACTTCGCGAGCGGAGAAACAGAG 3213
      |||||
987 erProAlaAlaAsnThrAlaSerGlnAlaGlnLysAlaThrGlnThrAsp 1003
3214 GCGCGCTTCGAGCGCAAA..... 3231
      |||||
1004 GtYAlaGlnIleAlaLysPheGlnAsnIleValAlaIleProPheSerPr 1020
3232 .....CAGGCAACAATTCCGCGCAACACAGCGCGGAAAGACA 3271
      |||||
1020 oGlnAlaAsnGlnAlaGlnLysGlnAlaLeuArgGlnGlnAlaLysAlaGlnG 1037
3272 AGCGG.....CAAGCGCTTACGCGCGTATGCGCGCGGCGGCGCAT 3312
      |||||
1037 InValLysArgGlnGlnAlaAlaGlnAlaGlnLysValAlaIleArgGlnLys 1053
3313 GCGACGCAAAAGGACAGAAAGTTCGCGCAACCGCGCGCGGACG..... 3354
      |||||
1054 AspArgGlnAlaLysArgLysAlaAlaGlnIleAlaArgGlnGlnGlnGln 1070
3355 .....GCGCGCGGCAAAATCCGCGCATTTATGCAAGCGCGGAGAGAGA 3397
      |||||
1070 uAlaArgLysAlaAlaGlnIleuAlaAlaLysGlnLysAlaGlnAlaGlnAla 1087
3398 AAAAAGCGGTGACGCGGATTAAGACACCGCGTTCGCGAAACAGCGGCA 3447
      |||||
1087 rGlyLysAlaArgLys.....LeuAlaArgGlnLysAla 1097
3448 GCGGAAACCGCGCGCTACACCGCTTCCCGCGCGCGCGCGCGCGCG 3497
      |||||
1098 GlnGlnAlaSerHisGlnAlaAsnAlaLysPheTyrArgArgArgArgAr 1114
3498 CCGGGAATTTGCGGCAACG..... 3516

```

```

      |||||
1114 gAlaIleLeuProAlaArgProAlaProValPheSerLeuAspArgTyrAla 1131
3516 ..... 3516
1131 sPrAlaLysAspAsnSerGlnSerIleGlyAsnLeuAlaArgValIle 1147
3516 ..... 3516
1148 rProArgMetGlyArgGlnLeuIleAsnArgTyrGlnGlnIleProLeuGln 1164
3516 ..... 3516
1164 uGlnLeuGlnAspGlnAlaGlnGlnGlnGlnArgGlnAlaThrGlnPheH 1181
3516 ..... 3516
1181 IsSerLysSerArgAsnArgArgAlaIleSerSerGlnProSerSerAsp 1197
3516 ..... 3516
1198 GlnAspAlaSerGlnSerValSerThrSerAspLysHisProGlnAspAs 1214
3516 ..... 3516
1214 nThrGlnLeuHisGlnLysValGlnIleThrAlaGlnLysGlnProArgAla 1231
3517 ..CAGCGCGACCGGCAACCGGCAACCGGCGGCGGCGGCGGCGGCTAT 3564
      |||||
1231 LeGlnProArgGlnGlnAlaAlaGlnAlaAspAlaValSerThrArgAsn 1247
3565 GCGCAATAGCGGTTTGATGATTTTCGCGACGCGCTTCACAGCGGTT.... 3609
      |||||
1248 ThrAsnSerIleAspSerArgAlaMetAlaSerThrGlnSerIleLeuLe 1264
3610 .....TTCGCGCTACAGACGCAATTTGACCGCGCGTTCGCGG 3646
      |||||
1264 uAspThrGlnLysAlaTyrLeuThrArgHisIleAlaGlnLysSerArgAla 1281
3647 AAGACGCGCGGCAACCGGCTTTCGACAAAGCGCATCCGCGGACACAAAC 3696
      |||||
1281 sPrAlaGlnLysAsnSerValTyrMetSerAsnThrGlyArgLysArgAsp 1297
3697 TACCGTTCGCAAGATTCGCGCGCTACCGCGCAACAAACGACGCGCGCA 3746
1298 TyrAlaSerAlaGlnIleTyrArgArgPheSerSerLysArgThrGlnThrGln 1314
3747 AATCGGATGCGCAAAACCTTCGCGACGCGCATCGCAACTCG 3793
      |||||
1314 nIleGlyIleAspArgSerLeuSerGlnIleAsnMetGlnIleGlyGlyVal 1331
3794 TTTGCGCAACCGGACCGGCAACCTTCGCGACGCGCATCGCAACTCG 3843
      |||||
1331 euThrTyrSerAspSerGlnHisIleThrPheAspGlnAla...GlyGlyLys 1346
3844 GCGCGCTTCGCGCGGCGGCGGCTTTCGCGCATTCAGCGCATCGCGG...AG 3890
1347 AsnThrPheValGlnAlaAsnLeuTyrGlnLysTyrTyrLeuHisnSprAl 1363
3891 GTTCGACATCGGCACTACGCGCGGCGGCGGCTTTAGTACGCGGACGCTTT 3940
      |||||
1363 atTrpTyrValAlaGlnLysArgIleGlyAlaGlnLysLeuArgSerArgLeuG 1380
1380 InThrGlnGlnLysAlaAsnPheAsnArgThrSerIleGlnThrGlnLys 1396
3991 CAGCGCAAGATACCGCGGCTTCGCGGATTCGCGCATCGCAACGCGCAT 4040
      |||||
1397 ThrLeuGlnAsnThrLeuLysIleAsnGlnPheGlnIleValProSerAl 1413
4041 CCGCGCAACGCGCTATTCGTCGCAAAAGCGGATTCGCGCAAAACG 4090
      |||||

```

```

1413 aGlylleaRgTyrSerArgLeuSerSerAlaAspTyrLysLeuGlyAspa 1430
4091 TCAATATGCGCACCCCGGCTTGCATTCAACCGCTACCCGGCGCAT 4140
1430 spSerValLysValSerSerMetAlaValLysThrLeuThrAlaGlyLeu 1446
4141 AAGGACATTTATTCATTCAACCGCGCACACATTCATTCACGCTTA 4190
1447 AspPheAlaTyrArgPheLys...ValGlyAsnLeuThrValLysProle 1462
4191 TTGAGCGCTGTCCTATACCATCGCCGCTCCGCAAGTCGACGCGG 4240
1462 uLeuSerAlaLalTyrPhe...AlaAsnTyrGlyLysGlyValAsn 1478
4241 TCAATACCGCGCTATTCGCGCAGGATTCGCAAAACCGCAGTGGGAA 4290
1478 alGlyGlyLysSerPheAlaTyrLysAlaAspAsnGlnGlnTyrSer 1494
4291 TGGGCGCTAAACCGCGCAATCAAAAGTTTCACGCTGCTCCACGCTG 4340
1495 AlaAlaGlyAlaLeuLeuTyrArgAsnValThrLeuAsnValAsn 1511
4341 CGCGCCGCAAGGCGGCAATTCGAGCGCAGCAGCAGCGCGCATCAAT 4390
1511 rIleThrLysGlyLysGlnLeuGlnLysGlnLysSerGlyGlnIleLys 1528
4391 TAGGTCACGCGCTGG 4404
1528 leGlnIleArgPhe 1532

seq_name: SwissProt_40: IGA4_HAEIN

seq_documentation_block:
ID IGA4_HAEIN STANDARD: PRT: 1849 AA.
AC P45386;
DR 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE Immunoglobulin A1 protease precursor (EC 3.4.21.72) (IGA1 protease).
GN IGA.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_Taxid=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NH71 HK61;
RX MEDLINE=92234949; PubMed=1373717;
RA Poulsen K., Reinholdt J., Kilian M.;
RT "A comparative genetic study of serologically distinct Haemophilus
influenzae type 1 immunoglobulin A1 proteases.";
RL J. Bacteriol. 174:2913-2921(1992).
CC -1- FUNCTION: VIRULENCE FACTOR; CLEAVES HOST IMMUNOGLOBULIN A
PRODUCING INTACT FC AND FAB FRAGMENTS.
CC -1- CATALYTIC ACTIVITY: Cleavage of immunoglobulin a molecules at
certain Pro-I-Xaa bonds in the hinge region. No small molecule
substrates are known.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- DOMAIN: THE SIGNAL PEPTIDE GUIDES THE PRECURSOR TO THE PERIPLASMIC
SPACE, AND THE CARBOXY-TERMINAL HELPER DOMAIN ASSOCIATES WITH THE
OUTER MEMBRANE TO FORM A PORE FOR EXCRETION OF THE PROTEASE
DOMAIN. THE HELPER DOMAIN IS THEN RELEASED BY AUTOPROTEOLYSIS (BY
SIMILARITY).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S6 (SERINE PROTEASE).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isdb.ch/announce/
CC or send an email to license@isdb.ch).

```

```

DR EMBL: M87491; AAA24968.1; -.
DR MEROPS: S06.001; -.
DR InterPro: IPR000710; IGA_S6.
DR Pfam: PF02395; IGA1; 1
DR PRINTS: PRO0921; IGASERPTASE.
KW Hydrolyase; Serine protease; Transmembrane; Zymogen; Signal.
FT SIGNAL 1 25
FT CHAIN 26 1021 POTENTIAL.
FT PROPEP 1022 1849 IMMUNOGLOBULIN A1 PROTEASE.
FT ACT_SITE 299 299 HELPER PEPTIDE (POTENTIAL).
SQ SEQUENCE 1849 AA; 202957 MW; 79A7D018C7150A6A CRC64;

```

```

alignment_scores:
    quality: 1516.50      length: 1917
    ratio: 1.523          gaps: 60
    Percent Similarity: 51.956      Percent Identity: 25.456

```

alignment_block:

US-09-303-518D-653 x IGA4_HAEIN ..

Align seg 1/1 to: IGA4_HAEIN from: 1 to: 1849

```

64 CGCTTCGCGCGCTTACTAGCATATGCGTGTGCGCATTTGCGC 113
   :::::
5 LysPheLysLeuAsnPheIleAlaLeuThrValAlaTyrAlaLeuThrPr 21
114 CCAAGCGCGCGCGGACACACTATTTCGCGCATCAACTCAATGCTATG 163
   I :::::
21 oTyrThrGluAlaIleValAlaArgAspArgValAspTyrGlnIlePheA 38
164 GCGACTTTGCCGAAATTAAGCAAGTTTGCAGTCGCGGCGGCAAGATAT 213
   I :::::
38 rGAspPheAlaGlnLysAsnLysGlyLysPheSerValGlyAlaThrAsnVal 54
214 GAGCTTTACAAACAAAGAGGAGTTCGCGCAATGATGATACCAAGC 263
   I :::::
55 GluValAlaArgAspLysLysAsnGlnSerLeuGlySerAlaLeuProAsnG 71
264 C...CGATGATGATTTTGTGTGATGCG...CGTACGCGCGTGGCG 307
   I :::::
71 yIleProMetIleAspPheSerValValAlaValAspLysArgIleAlar 88
308 CATTCGCGCGCGCATGATATTTGAGCGCGCAT...AAGCGCGC 354
   :::::
88 hIleuValAsnProGlnTyrValAlaGlyValLysHisValSerAsnGly 104
355 TATAACATGTTGATTTGCT.....GCGAGGAGGAGCA 389
   :::::
105 ValSerGlnLeuHisPheGlnLysLeuAsnGlnLysAsnMetAsnAsnGlyAs 121
390 TCCCGATCGACGCGCG.....TTTCTTACCAAAATG 421
   :::::
121 hAlaLysSerHisArgAspValSerSerGlnGlnAsnArgTyrTyrThrV 138
422 TGAAGAAGATATATAT.....AAA 441
   I :::::
138 alGluLysAsnSnPheProThrGlnAsnValThrSerPheThrLys 154
442 GAGGAGACTAACGCGCATCTTAGCGCGGATTCATGATGATGCGCGCTT 491
   :::::
155 GluGlnGlnAspAlaGlnLysArgArgLysAspTyrTyrMetProAlaGle 171
492 GCACAAATTTGCGACAGATGCGAACCCTGTGAG...ATGACAGATTATA 538
   I :::::
171 uAspLysPheValThrGlnValAlaProIleGlnAlaSerThrAlaAsnA 188
539 TGGATGGGTGGAATATACCTGATTTAAATAAATGCTGATCGTGTGCA 588
   :::::
188 snAsnLysGlyGlnTyrAsnAsnSerAspLysTyrProAlaPheValArg 204
589 ATCGAGACGCGCAGCAATAT..... 609
   :::::

```

```
205 LeuGlySerGlySerGlnPheIleTyrLysGlySerArgTyrGlnLeu 221
610 .TGGCGGTCTGATGAGACGAACCCATTAACCGCGAAGTTCATATCAT. 657
221 uIleLeuThrGlnLysAspLysGlnGlnLysnLeuLeuArgAsnTrpAspV 238
658 .....ATTGCAAGCCGATATTCCTGGCTGCTC 684
238 aLgLyGlyAspAsnLeuGlnLeuValGlyAsnAlaTyrThrTyrGlyIle 254
685 GGGGGCAATACCTTTCACAAATGGATGAGTGGGGGACAGACAGTCACT 734
255 AlagLyThrProTyrLysValAsnHisGlnAsnGlnLysnLeuLeuLpH 271
735 AGGTAGCAAAAAATTAAACATAGC...CCATATGGTTTTTTTACCAACAG 781
271 eGlyAsnSerLysGlnGlnLysSerAspProLysGlyIleLeuSerGlnA 288
782 GA.....GGCTCATTTGGCGACAGTGGCTCACCAGATGTTT 816
288 sPProLeuThrAsnTyrAlaValLeuLysAspSerGlySerProLeuPhe 304
817 ATCTATGATGCCAAAGCAAAAGTGGTATTATTAATGGGGTATTGCAAC 866
305 ValTyrAspArgLysGlyLysTyrLysTrpLeuPheLysGlySerTyrAspH 321
867 AGGCACCCCTATATAGAAAAAGCAATGGCTTCACAGTTCGTAAAG 916
321 eTrpAlaGlyTyr.....AsnLysLysS 929
917 ATTGGTCTATGATGAATCTTTCGTGAGATACCATTCAGTATTCATAC 966
329 eTrpGlnGlnTrpAsnLysTyrLysHisGlnPheAlaGlnLysIleTyr 345
967 GAACCACTCAAAATGGAAATCTTTTAAAGCAATAATAATAGCGGC 1016
346 GlnGlnTyr.....SerAl 350
1017 AGGAAAAATCATGCCAAACATTAACATATCTCTACCTTATGATTA 1066
350 agLysSerLeuThrGlySerAsnThrGlnTyrThrGlnAlaThrGlyS 367
1067 AAACACGAAACGGT.....CAATGTTTAAATGTTCTTTTATCC 1104
367 eTrpHiserThrIleThrGlyGlyGlnProLeuSerValAspLeuThr 383
1105 GAGACAGCAAGAACCTGTTTATCATCTCTCAGTGGGGTCAACAGTGA 1154
384 AspGlyLysAspLysPro..... 389
1155 TCGACCCAGACGTGAATTAATGAGAAATATTTCTTATTTGACAAAGGA 1204
390 .....AsnHisGlyLysSerIleThrLeu.....LysGlyS 400
1205 AAGGTGAATGTACTTACACGAAACATCAACCAAGCGCGGGGGTGTG 1254
400 eGcLyThrLeuThrLeuAsnAsnHisIleAspGlnGlnLysIleGlyLeu 416
1255 TATTTTGAAGGTAAATTTTACGTC...TCGGCTAAAAACAAGCAAGAA 1301
417 PhePheGlnGlnLysPtyrGlnValLysGlyThrSerAspSerThrThr 433
1302 GCAAGCGCGGGCGTTCATATCAGTATGAGTACGCTACCTTACTTGAAG 1351
433 pLysGlyAlaGlyAlaSerValAlaAspGlyLysThrValThrTrpLysV 450
1352 TAAACGGCGTGGCAACGACGCGCTGTCCAAAATGGCAAGGCGCGTGG 1401
450 aLHisAsnProLysTyrAspArgLeuAlaLysIleGlyLysGlyThrLeu 466
1402 CTGGTTCAAGCCAAAGGGAACCAAGGCTCGGTACAGGTGGGGGACGG 1451
467 ValValGlnGlnLysGlyLysAsnGlnGlyLeuLeuLysValGlyAsp 483
1452 TAAAGTCACTTTAGATCAGACGGCGAGCATCAGCAAAACAAAGCCT 1501
483 yThrValIleLeuLysGlnLysAlaAspAlaAsnLysValGlnLap 500
1502 TTAGGAATCGGGCTGTGTCAGCGGACGGAGCGGTGCACTGAATGCC 1551
500 hSerGlnValGlyIleValSerGlyArgSerThrIleValLeuAsnAsp 516
1552 GATAATCAGTTCAACCCCGCAAACTATTTCGGCTTTGGCGGGGAGC 1601
517 AspLysGlnValAspProAsnSerIleTyrPheGlyPheArgGlyArg 533
1602 TTTGGATTTGAACGGGCATTCGCTTCGTTCACCGCATTCAAATACCG 1651
533 gLeuAspLeuAsnGlyAsnSerLeuThrPheAspHisIleLeuArgSnIleA 550
1652 ATGAAGGGCGCGATGTGTCAAACCAATCAAGCAAAAGAAATCCACGTT 1701
550 sPAspGlyAlaArgValAlaAsnHisAsnMetThrAsnThrSerAsnIle 566
1702 ACCATTACAGCGCAATAAGATATTACTACA.....ACGGCAA 1739
567 ThrIleThrGlyLysSerLeuThrAsnProAsnThrIleThrSerTyr 583
1740 TAAC..... 1743
583 rAsnIleGlnAlaGlnAspAspAspHisProLeuArgIleArgSerIleP 600
1743 ..... 1743
600 rOtyrArgGlnLeuTyrPheAsnGlnAspAsnArgSerTyrTyrThrLeu 616
1743 ..... 1743
617 LysLysGlyAlaSerThrArgSerGlnLeuProGlnAsnSerGlyLysE 633
1743 ..... 1743
633 rAsnGlnAsnTyrPLeuTyrMetGlyArgThrSerAspGlnAlaLysArgA 650
1744 .....AACAACTGGATAGCAAAAAGAAATGGCTTCACACGCTGG 1785
650 snValMetAsnHisIleAsnAsnGlnArgMetAsnLysPheAsnGlyTyr 666
1786 TTTGGCGAGAAAGATGCAACCAAAACG...AACGGCGGCTCAATCTGAA 1832
667 PheGlyGlnGlnGlnLysThrLysAlaThrGlnAsnGlyLysLeuAsnValTh 683
1833 TTTACCAACCGGAAGAAGCGGATCGCACTTACTGCTTCGGCGGAACA 1882
683 rPheAsnGlyLysSerAspGlnAsnArgPheLeuLeuThrGlyGlyThrA 700
1883 ATTTAAACGGCAATATACGCAAAACAAACGCAAACTGTTTTCACGGC 1932
700 snLeuAsnGlnLysPLeuAsnValGlnLysGlyThrLeuPheLeuSerGly 716
1933 AGACCGACCGGACCGCTTCAATCATTTAAGAACGGGGGTGCACAAAT 1982
717 ArgProThrProHisAlaArgAspIleAlaGlyIleSerSerThrLys 733
1983 GGAAGGT.....ATCCCAACAGAGAAATCGTGGGACCAACGATGGA 2026
733 sAspProHisPheThrGlnLysAsnGlnValValAlaGlnAspAspTyr 750
2027 TCGACCGCACATTTAAAGCGGAAACTTCATATTCAGCGGAGACAAGCG 2076
750 LAsnAlaGAsnPheLysAlaThrThrMetAsnValThrGlyAsnAlaSer 766
2077 GTGGTTTCC...CGCAATGTTGCCAAAGTGAAGGCGATTTGGCATTAAG 2123
767 LeuTyrSerGlyArgAsnValAlaAsnIleThrSerAsnIleThrAlaSe 783
```


2124 CAATCCGCCCAAGCAGTTTTCGGTGTGCGACCGCATCAAGCACACAA 2173
||||| :|||
783 rAsnAsnAlaGlnValHisIleGlyTyr.....LysThrGlyAspThrV 798
2174 TCGTACAGCTTCGGACGTGGAGGCTCTACAGATTGACGAAAAAAC 2223
:||||| :|||
798 aLcysValArgSerAspTyrThrGlyValThrCysHisAsnSerAsn 814
2224 ATTACGACGATTAAGTGTATGCTCATTTGACAGACAGACATGAGG 2273
:||||| :|||
815 Leuser...GluLysAlaLeuAsnSerPheAsnProThrAsnLeuArgI 830
2274 CAATGTACGCTTGGCGATACGCTCATTTAAATCTCAAGACATTTGCA 2323
||||| :|||
830 yAsnValAsnLeuThrGluAsnAlaSerPheThrLeu...GlyLysAla 846
2324 CACTCAAGCGCAATCTAGTCAGGCGA..... 2352
:||||| :|||
846 snLeuPheGlyThrIleGlnSerIleGlyThrSerGlnValAsnLeuLys 862
2353 ...GACAGCACTATACGTTACGGCAAGCGCACCC..... 2385
:||||| :|||
863 GluAsnSerHisThrHisLeuThrGlyAsnSerAsnValAsnGlnLeuAs 879
2386 ...CAAAAGCGCAACTCAGCCTCGTGGCAATGCCCA..... 2421
||||| :|||
879 nLeuThrAsnGlyHisIleHisLeu.....AsnAlaGlnAsnAspAla 894
2422GCACATTTAATCAAGCCACATTAAGCGCAACACATCGGCT 2463
:||||| :|||
894 snLysValThrThrTyrAsnThrLeuThrValAsn.....SerLeu 907
2464 TCGGACAAATGCTTCATT.....AATCTAAG 2489
||| :|||
908 SerGlyAsnGlySerPheTyrTyrValAspPheThrAsnAsnLysSe 924
2490 CACACAGCCCGTACAAAGCGCAGT..... 2514
||||| :|||
924 rAsnLysValValValAsnLysSerAlaThrGlyAsnPheThrLeuGlnV 941
2515CTGACGCTTTCCGAC 2529
:||||| :|||
941 aLAlaAspLysThrGlyLupProAsnHisAsnGlnLeuThrLeuPheAsp 957
2530 AACGCTAAGCA.....AACGTAAGCATTCGCGCATCAACGCGCA 2570
:||||| :|||
958 AlAserAsnAlaThrArgAsnAsnLeuGlnValThrLeuAlaAsnGlySe 974
2571 TGTCTCCCTAGCCGATTAAGCAATTCATTGAAAAACAGCGCTTTA 2620
:||||| :|||
974 rVal.....AspArgGlyAlaTyrLysTyrLysLeuArgAsnValA 988
2621 CCGGAAAATGACGGCGGCAAGATTAAGCATTAACCTTAAAGACAGC 2670
:||||| :|||
988 snGlyArgTyrAsp.....LeuTyrAsnPro 996
2671 GAATGACGCTGCCGTCGGACGCAATTAAGCAATTTAACTTGACAA 2720
||| :|||
997 Glu.....ValGlnLysArgAsnGlnThrValAspTh 1007
2721 CGGCACATTAAGCTATCCGCTATCGACAGATGCGGACGCGCGC 2770
:||||| :|||
1007 rThrAsnIleThrThrProAsnAspIleGlnAlaAspAlaProSerAlaG 1024
2771 AAACGCGCAGT..... 2781
:||||| :|||
1024 InsrAsnAsnGlnGlnIleAlaIleArgValGlnThrProValProProPro 1040
2782GCGGACAGTCCGCGCGCGCGCT. 2805
:||||| :|||
1041 AlAlaProAlaThrGlnSerAlaIleAlaSerGlnProGlnThrArgPr 1057
2805 2805

1057 cAlaGlnThrAlaGlnProAlaMetGlnGlnThrAsnThrAlaAsnSer 1074
2806TCG 2808
1074 hrGlnThrAlaProLysSerAspThrAlaThrGlnThrGluAsnProAsn 1090
2809 CGCGCTCCCTATATCCGTTACG..... 2832
:||||| :|||
1091 SerGlnSerValProSerGlnThrThrGlnLysValAlaGlnAsnProPr 1107
2832 2832
1107 oGlnGluAsnGlnThrValAlaLysAsnGlnGlnAlaThrGlnProT 1124
2833CGCCCACTTCGCA 2847
:||||| :|||
1124 hrProGlnAsnGlyGluValAlaLysGlnAspGlnProThrValGluAla 1140
2848 GAATCCGGTTTCACACAGCTGACGCTAAGC...GCCAATTAAGAGGTCA 2894
:||||| :|||
1141 AsnThrGlnThrAsnGlnAlaThrGlnSerGlnGlyLysThrGlnGln 1157
2895 GGGACATTCGCTTATGTCGAA..... 2919
||| :|||
1157 rGlnThrAlaGlnThrLysSerGlnProThrGlnSerValThrValSerG 1174
2920CTGCTTC 2925
1174 LuAsnGlnProGlnLysThrValSerGlnSerThrGluAspLysValVal 1190
2926 GCCTACCGGACGCGCAATTAAGCTG..... 2952
:||||| :|||
1191 ValGlnLysGlnGlnLysAlaLysValGlnThrGlnGlnThrGlnLysAl 1207
2952 2952
1207 aProGlnValThrSerLysGlnProProLysGlnAlaGlnProAlaProG 1224
2953CGGAAAGTTCGAAAGC 2970
:||||| :|||
1224 LglLupValProThrAspThrAsnAlaGlnGlnAlaLeuGlnGln 1240
2971 ACTTACACCTTGCTGTC..... 2988
||| :|||
1241 ThrGlnProThrThrValAlaAlaAlaGlnThrThrSerProAsnSerLy 1257
2988 2988
1257 sProAlaGlnGlnThrGlnGlnProSerGlnLysThrAsnAlaGlnProV 1274
2989AACATACCGGACAGACACCCGTAAGTCTCGAG 3021
:||||| :|||
1274 aThrProValValSerGluAsnThrAlaThrGlnProThrGlnThrGln 1290
3022 CAATTGACGCTAGTGAAGAAAGACACACACCCGCTGCGAAATCT 3071
:||||| :|||
1291 GluThrAlaLysValGlnLysGlnLysThrGlnGlnValProGlnVal. 1306
3072 TAAATTCACCTTCGCAACACACAGCAGTGCAGGCGGCGATATC 3121
:||||| :|||
1307AlaSerGlnLysSerProLysGlnGlnProAl 1318
3122 AGCTTATCCGCAAGACGGGAGTTCGCGCTGATATACCGGCAAGAA 3171
:||||| :|||
1318 aAlaLysProGlnAlaGlnThrLysProGlnAlaGlnProAlaArgGln 1334
3172 CAAGACCTTTCGACAA..... 3189
:||||| :|||
1335 AsnValLeuThrThrLysAsnValGlyGlnProGlnProGlnAlaGlnPr 1351
3190CTGCGGACGCGGAGAGAAACAGAGCC...G 3217
:||||| :|||

RA Yamamoto Y., Alba H., Baba T., Hayashi K., Inada T., Isono K.,
 RA Itoh T., Kimura S., Kitagawa M., Makino K., Miki T., Mitsuhashi N.,
 RA Mizobuchi K., Mori H., Nakade S., Nakamura Y., Nashimoto H.,
 RA Oshima T., Oyama S., Saito N., Sempel G., Satoh Y., Sivasubram S.,
 RA Tagami H., Takahashi H., Takeda J., Takemoto K., Uehara K., Wada C.,
 RA Yamagata S., Horikuchi T.,
 RT Construction of a contiguous 874-kb sequence of the *Escherichia coli*
 RT - K12 genome corresponding to 50,068.8 min on the linkage map and
 RT analysis of its sequence features.";
 RL DNA Res. 4:91-113(1997).
 CC - SUBCELLULAR LOCATION: Outer membrane (Potential).
 CC - SIMILARITY: STRONG, TO BORDETELLA PERTACTIN.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 DR EMBL: U36840; AAA79815.1; ALT_SEQ.
 DR EMBL: AE000350; AAC75695.1;
 DR EMBL: D90889; BAA16514.1; ALT_INT.
 DR EMBL: D90890; BAA16518.1; ALT_INT.
 DR Ecogene: EG13213; ypfA.
 KW Hypothetical protein; Outer membrane; Complete proteome.
 KW SEQUENCE 1569 AA; 162774 MW; 0D006A9A32154596 CRC64;

alignment_scores:
 Quality: 359.00 Length: 1686
 Ratio: 0.481 Gaps: 80
 Percent similarity: 44.306 Percent Identity: 19.395

alignment block:

US-09-303-518D-653 x YPJA_ECOLI ..

Align seg 1/1 to: YPJA_ECOLI from: 1 to: 1569

```

445 GGGACTAAGCGCCATCTTATGGCGGCGATATCATATGCGCGTTTGA 494
      |||||.....:..:.....:..:.....:..:.....:..:
97 G|YThrThrasnsthrlstleleasnhslyglc|nlnsnvalh 113
      |||||.....:..:.....:..:.....:..:.....:..:
495 CAATTTTGCACAGATGACAGAACCTTTAGATGACACCTTATGATG 544
      |||||.....:..:.....:..:.....:..:.....:..:
113 sgllyglvalserasnlyserleuilegluserglytyrlylnaspi 130
      |||||.....:..:.....:..:.....:..:.....:..:
545 GGTGAATACGCTGATTAATAATACCCTGATCGTTCGATCGGA 594
      |||||.....:..:.....:..:.....:..:.....:..:
130 leglserhsasnsthnevalglc|lnlnlnasnsthrlstleasn 146
      |||||.....:..:.....:..:.....:..:.....:..:
595 GCAGGACGACATATTGGCGGCTGTGATGAAGACCAACCAATACCGCA 644
      |||||.....:..:.....:..:.....:..:.....:..:
147 G|YglYlAArgln.....:..:.....:..:.....:..:.....: 150
      |||||.....:..:.....:..:.....:..:.....:..:
645 AATTTATATCATATTGCA.....AGCCATATTCTTGCGTCGCGTG 688
      |||||.....:..:.....:..:.....:..:.....:..:
151 ....Serllehsasp|lyglYlleserthrlYthrllegluserg 166
      |||||.....:..:.....:..:.....:..:.....:..:
689 GCATATCTTTGACAAATGGA.....TCAGGTGGTGCACAGACGCAAC 732
      |||||.....:..:.....:..:.....:..:.....:..:
166 lYasnlnaspvalYlrylslyglYlleaserasnelythrlstlely 182
      |||||.....:..:.....:..:.....:..:.....:..:
733 TTAGGTAGGCAAAATTAATTA.....CATAGCCCATATGGTTTTCAC 779
      |||||.....:..:.....:..:.....:~:.....:~:
183 glYglYlA|aserArYvalglc|lyserAlaasnlylleleuileas 199
      |||||.....:..:.....:..:.....:~:.....:~:
780 AGAGGCTCA.....:~:.....:~:.....:~:.....:~:.....T 790
      |||||.....:~:.....:~:.....:~:.....:~:.....T
199 pglYglYsergl|nllvalYlvalglc|nlyhlsAlaasp|lythrl 216
      |||||.....:~:.....:~:.....:~:.....:~:.....T
791 TTGGCAGACAGTGCACCAATGTTTATCTATGATGCCCAAAAGCAAG 840
      |||||.....:~:.....:~:.....:~:.....:~:.....T

```

```

216 leasnlyserglYsergl|naspvalYlglc|lyserleuAlaThrAsn 232
      |||||.....:~:.....:~:.....:~:.....:~:.....
841 TGGTTATTAATGGGTATTCGAACAGGCAACCCCTTATGCA..... 885
      |||||.....:~:.....:~:.....:~:.....:~:.....
233 ThrThrlleasnly.....:~:.....:~:.....:~:.....:~: 245
      |||||.....:~:.....:~:.....:~:.....:~:.....
886 .....:~:.....:~:.....:~:.....:~:.....:~:..... 913
      |||||.....:~:.....:~:.....:~:.....:~:.....
245 lThrYalgluThrlThrlleYsa|nlyglc|lyglu|nArYalYrG 262
      |||||.....:~:.....:~:.....:~:.....:~:.....
914 AAGATTGTTTCATATGATGAATCTTCTCGAGATACCATTCATATTC 963
      |||||.....:~:.....:~:.....:~:.....:~:.....
262 lserArYalAleu|s|ph|Thrlleglu|glYlThrlGlnSerleu... 277
      |||||.....:~:.....:~:.....:~:.....:~:.....
964 TACGAACCATCAATCAAAATGGAAATACTTTTAAAGCAATATATAGG 1013
      |||||.....:~:.....:~:.....:~:.....:~:.....
277 .....:~:.....:~:.....:~:.....:~:.....:~:..... 277
      |||||.....:~:.....:~:.....:~:.....:~:.....
1014 CGCAGGAAATCGATGCGCAACATTAACACTATTCCTTACCTTATGAT 1063
      |||||.....:~:.....:~:.....:~:.....:~:.....
278 AsnSerlySerThrlAla|Ysa|nThrlHlsle.....:~:~:~: 291
      |||||.....:~:.....:~:.....:~:.....:~:.....
1064 TAAAAACAGACCGCTCAATGTTTAAATGTTTCTTATTCGAGACAGCA 1113
      |||||.....:~:.....:~:.....:~:.....:~:.....
291 lYglYlThrlc|lnlevalasp.....:~:~:~:SerThrSerAspvalle 305
      |||||.....:~:.....:~:.....:~:.....:~:.....
1114 AGAGAACCTGTTTATATGCTGCAGGTGGGTCAACAGTTATCGACCCG 1163
      |||||.....:~:.....:~:.....:~:.....:~:.....
306 Glu.....:~:~:~:Valtyr.....:~:~:~:Ser|glYlYalAleuAspValArg... 316
      |||||.....:~:.....:~:.....:~:.....:~:.....
1164 ACTGAATATGAGAAATATTTCTCTTATATGACAAAGCAAAAGTGAT 1213
      |||||.....:~:.....:~:.....:~:.....:~:.....
317 .....:~:~:~:GlYglYlThrlAla|ThrAsnValThrlHlsasp|YalAl 330
      |||||.....:~:.....:~:.....:~:.....:~:.....
1214 TGATCTTACAGCAACATCAACCAAGCGCGGCTTGTATTTGAG 1263
      |||||.....:~:.....:~:.....:~:.....:~:.....
330 leleu|YstThrasnThrasn|lyThrlThrValserglYlThrasSer|lu 346
      |||||.....:~:.....:~:.....:~:.....:~:.....
1264 GGTATTTTACAGGTCTCCCTTAAAAACAGCAACAGTGGCAGCGCGG 1313
      |||||.....:~:.....:~:.....:~:.....:~:.....
347 GlYAla|PheSerle.....:~:~:~:HlsasnHlsal.....:~:~:~:AlaAspAs 358
      |||||.....:~:.....:~:.....:~:.....:~:.....
1314 GCTTATATCAATGATGACAGTACCGTACTTGAAGTAAACGCGGTG 1363
      |||||.....:~:.....:~:.....:~:.....:~:.....
358 nValleu|eu|uasn|lyglYlHlsleuAsp|lleasn|Ala|Yr|glYserA 375
      |||||.....:~:.....:~:.....:~:.....:~:.....
1364 CAACGACCGCTGTCAAAATCGGCAAAAGCGACGCTG.....CTGGTT 1407
      |||||.....:~:.....:~:.....:~:.....:~:.....
375 laasn|YstThrllelely.....:~:~:~:Asp|Ysg|YlThrMetSerValleuThr 390
      |||||.....:~:.....:~:.....:~:.....:~:.....
1408 CAAGCCAAAGG.....:~:~:~:GAACCAAGCTTCGTCACCGT 1442
      |||||.....:~:.....:~:.....:~:.....:~:.....
391 AsnAla|Ysa|Ala|spAla|ThrAr|lleasn|YglYlAla|MetAspVa 407
      |||||.....:~:.....:~:.....:~:.....:~:.....
1443 GCGCGACGCT.....:~:~:~:AAAGTCATCTTATGATCAGCAGCGGACGATCAAG 1486
      |||||.....:~:.....:~:.....:~:.....:~:.....
407 lAla|glYsa|nAla|ThrasnThrlleleasn.....:~:~:~:G 418
      |||||.....:~:.....:~:.....:~:.....:~:.....
1487 GCATTAACCAAGCCTTATAGTAAATCGGCTTGTCACGCGC..... 1527
      |||||.....:~:.....:~:.....:~:.....:~:.....
418 lYglYlThrlGlnasn|lleasn|nTyrg|lyllealAr|n|glYlThrasn|le 434
      |||||.....:~:.....:~:.....:~:.....:~:.....
1528 ...AGGGGACGCGTGCACATGAATGCGGATATAGTTCAACCCGCA 1574
      |||||.....:~:.....:~:.....:~:.....:~:.....
435 AsnserglYlThrlGlnasn|lleasn|Yser|glYlYsa|Ala|sp|Thrl 451
      |||||.....:~:.....:~:.....:~:.....:~:.....
1575 ACTCTATTTGCGCTTCGC.....:~:~:~:G 1594
      |||||.....:~:.....:~:.....:~:.....:~:.....
451 e|le|serSer|YserAr|rglnVal|glu|Ysa|sp|glYlThrAla|leG 468
      |||||.....:~:.....:~:.....:~:.....:~:.....
1595 GCGGACGTTGATTAACGCGCATTCGCTTCGCTTC.....:~:~:~: 1632
      |||||.....:~:.....:~:.....:~:.....:~:.....

```

468 lysAsnIleSerAlaGlySerLeuIleValTyrThrGlyGlyLe 484
 1633 ..CACCGCATTCAAAATACGATGACAGGGCGATGATTCAC... 1674
 485 AlaHisGlyValAsnGlnGlnThrGlySerAlaLeuValAlaAsnThrG1 501
 1675CACAAATCAAGACAAAGAAATCCACCGTTACCATTA 1708
 501 yAlaGlyThrAspIleGlnGlyTyrAsnLysLeuSerHisPheThrIleT 518
 1709 CAGCAATATAA.....GATTTACT 1728
 518 hrGlyGlyGluAlaAsnTyrValAlaLeuGluAsnThrGlyGluLeuThr 534
 1729ACAAACGGCGCATACAAACAACTTGATACGAAAAAAGA 1766
 535 ValValAlaLysThrSerAlaLysAsnThrIleAspThrGlyGlyLys 551
 1767 AATTGCTTACAAACGGTGTGGCGAGAAAGATCAACCAAAAGCAAG 1816
 551 sIeuIleVal.....GlnLysGluAlaLysThrAspSert 563
 1817 GCGGCTCAT.....CTG 1830
 563 hrArgLeuAsnAsnGlyGlyValLeuGluValGlnAspGlyGlyGluAla 579
 1831 AATTGCCAACCGGAGAGCGGATGCGACTTACTGCTTCC..... 1872
 580 LysHisValGlnGlnIleSerGlyAlaLeuIleAlaSerThrThSe 596
 1873GCGCAACAAATTTAAACGCAAT..... 1896
 596 rGlyThrIleuIleGlnGlyThrAsnSertGlyAspAlaPheTyrIleA 613
 1896 1896
 613 rGAsnSerGluAlaLysAsnValValLeuGluAsnAlaGlySerLeuThr 629
 1897ATCAGCAACAAACGGCAA 1916
 630 ValValThrGlySerArgAlaValAspThrIleIleAsnAlaAsnGlyL 646
 1917 ACTGTTTTCACGGCAGACCGACACCGCCTACATCATTTAGGA 1966
 646 sMetAspValTyrGlyLysAspValGlyThrValLeuAsnSerAlaGlyT 663
 1967 GC..... 1968
 663 hrGlnThrIleTyrAlaSerAlaThrSerAspLysAlaAsnIleLysGly 679
 1969GGGTGTCAAAAAATGGAAGGTATCCACCAAG 2000
 680 GlyLysGlnThrValTyrGlyLeuAlaThrGluAlaAsnIleGluSerG1 696
 2001 AGAAATCGGTGGGACACAGATTGGATCGACCGCATTTAAAGCGGAA 2050
 696 yGlnGlnIleValAspGlyLysThrGluLysThr..... 708
 2051 ACTTCATATTCAGGCGCA..... 2070
 709HisIleAsnGlyLysThrGlnThrValGlnAsnTyrGlyLysAla 723
 2071CAAGCGGTGTTCCCGCAATGT 2093
 724 IleAsnThrAspIleValSerGlyLeuGlnIleMetAlaAsnGlyLys 740
 2094 TGCCAA.....GTGGAAGCGATGGCATTTAAGCATCACGCC 2134
 740 rAlaGlnGlySerIleIleAsnGlyLysSerGlnValValAsnGlnGlyG 757
 2135 AAGCAGTTTTCGGTGCACCGCATCAAGCAACGACATCTGTACAGCT 2184
 757 LysLeuAlaGluAsnSerValLeuAsnAspGlyLysThrLeuAspValArg 773

2185TCGAGCTGACGGGCTCG..... 2202
 774 GlnLysGlySerAlaThrGlyIleGlnGlnSerSerGlnGlyAlaLeuVal 790
 2203ACAGTTTACGAAAAAACCATTTACCGACGATTAAGTATG 2245
 790 lAlaThrThrArgAlaThrArgValThrGlyThrArgAlaAspGlyValA 807
 2246 CTTCATTGAGCAGACCGACATCAGAGCAATGTACGCTTGGCGATCAC 2295
 807 lApeSerIleGlnGlnGlyAlaAlaAsnIleLeuLeuAlaAsn... 822
 2296 GCTCATTTAAATCTCAAGGACTTGCACATCAACGGAATCTAATGTC 2345
 823GlyGlyValLeuThrValGluSerAspThrSerSe 834
 2346 A.....GCGGAGACACGCACTATACGTTACCGC 2374
 834 rAspLysThrGlnValAsnMetGlyGly.....ArgGluIleValLysT 849
 2375 GCAACGCCACCCAAACGCGACCTCAGCCTCGGCAATGCCAA... 2421
 849 hrLysAlaThrAlaThrGlyThr...ThrLeuThrGlyGlnGlnIle 864
 2422GCAACATTTAATCAAGCCACATTAAC.....GCGACACATC 2459
 865 ValGlnGlyValAlaAsnGlnThrThrIleAsnAspGlyGlyIleGlnTh 881
 2460 GCGTTCGGACAAATGCTTCATTTATCTAAGCACAACGCGGTCAAAAG 2509
 881 rValSerAlaAsnGlyGlu...AlaIleLysThrLysIleAsnGlnGlyG 897
 2510 GCAGCTGACGCTTCGACACACGCTAAGCAAC.....GTAAACCAT 2553
 897 LysThrLeuThrValAsnAspAsnGlyLysAlaThrAspIleValGlnAs 913
 914 SerGlyAlaAlaLeuGlnThrSerThrAlaAsnGlyIleGluIleSerG1 930
 2604 TGAANAACGCCGCTTACCGGAAAAATCAGCGGCAACGATAGCGCAT 2653
 930 yThrHisGlnTyrGlyThrPheSerIleSerGlyAsnLeuAlaThrAsn 947
 2654 TACACTTAAAGACAGC..... 2670
 947 eLLeuLeuGluAsnGlyGlyAsnLeuValLeuAlaGlyThrGluAla 963
 2671 ...CAATGGACGCTGCCGTGGGACGGAATTAGCATTTAAACCTTGA 2717
 964 ArgAspSerThrValGlyLysGlyAlaMetGlnAsnLeuGlnAs 980
 2718 CAAGCCACCACTTACACTCAATTCGCGCTATCGACAC..... 2754
 980 pSerAlaThr...LysValAsnSerGlyGlnTyrThrLeuGlyArgS 996
 2755GATCGCGCA 2763
 996 eLysAspGluPheGlnAlaLeuAlaArgAlaGluAspLeuGlnValAla 1012
 2764 GGC.....GGCAAAACGGGACGCGGACATGCGCGCGCGC 2801
 1013 GlyGlyThrAlaIleValTyrAlaGlyThrLeuAlaAspAlaSerValSe 1029
 2802 CCGTTCGCGCGCTTCCTATATCCGTTACGCC..... 2835
 1029 rGlyAlaThrGlySerLeuSerLeuMetThrProArgAspAsnValThrp 1046
 2835 2835
 1046 roValLysLeuGlnGlyAlaValAlaArgIleThrAspSerAlaThrLeuThr 1062


```

907 GTTCGTAAGATGTTCTATGATGAATCTTTGCTGAGATACCATTC 956
    : : : : : : : : : : : : : : : : : : : : : : : :
125 LeuThr ..... AspValIlePheSerGlyAsnValAlaG1 136
    : : : : : : : : : : : : : : : : : : : : : : : :
957 AGTATTCTACGACACCATCAAAATGGAAATACCTTTTAAAGACATA 1006
    : : : : : : : : : : : : : : : : : : : : : : : :
136 yGlyTyr.....GlyGlyAlaIleTyrSerSerGlyT 147
    : : : : : : : : : : : : : : : : : : : : : : : :
1007 ATATGGCGGAGAAATATGATGCCAAACATTAACACTTTCCTACCT 1056
    : : : : : : : : : : : : : : : : : : : : : : : :
147 hrAsnAspThrGlyAlaValAspLeuArg..... 156
    : : : : : : : : : : : : : : : : : : : : : : : :
1057 TATAGATTTAAAAACACGACCGTTCAATGTTTAAATGTTCTTATCCG 1106
    : : : : : : : : : : : : : : : : : : : : : : : :
157 .....ValThrAsnAlaMetPheArgAsnAsnIleAlas 168
    : : : : : : : : : : : : : : : : : : : : : : : :
1107 GACACGACAGAACCTGTTTATCATGCTGACGGGCGGTCAAGATTATC 1156
    : : : : : : : : : : : : : : : : : : : : : : : :
168 naSpGlyLys.....GlyGlyAlaIleTyrThr.... 177
    : : : : : : : : : : : : : : : : : : : : : : : :
1157 GACCCAGACTGAATTAATGAGAAATATTTCTTATTGACAAAGAAAA 1206
    : : : : : : : : : : : : : : : : : : : : : : : :
178 .....IleAsnAspValTyrLeuSerAspValIlePheAspAsn 191
    : : : : : : : : : : : : : : : : : : : : : : : :
1207 GGTGAATTGATCTTACCTACGACCAATCAACCAAGCGCGCGGTTTG.. 1254
    : : : : : : : : : : : : : : : : : : : : : : : :
192 AsnGlnAlaIleTyrThrSerThrSerTyrSerAspGlyAspGlyAlaI 208
    : : : : : : : : : : : : : : : : : : : : : : : :
1255 .....TATTGTCG 1261
    : : : : : : : : : : : : : : : : : : : : : : : :
208 eaSpValIThrAspAsnAsnSerAspSerLysHisProSerGlyTyrThrI 225
    : : : : : : : : : : : : : : : : : : : : : : : :
1262 AGGGTAAATTTTACGCTCTCGCCTAAACACGAAACGTCGCAAGCGCG 1311
    : : : : : : : : : : : : : : : : : : : : : : : :
225 LeValAsnAsnThrAlaPheThrAsnAsnThrAlaGlnGlyTyrGlyGly 241
    : : : : : : : : : : : : : : : : : : : : : : : :
1312 GGGGTCATATCAGTGTGACGAGTACCGTTACTTGAAAGTAAAGCGCGT 1361
    : : : : : : : : : : : : : : : : : : : : : : : :
242 AlaIleTyr...ThrAsnSerValThrAlaProTyrLeuIleAspIleSe 257
    : : : : : : : : : : : : : : : : : : : : : : : :
1362 GGCAAAGCGCGCTGTCCAAATGCGCAAGGACGACGCTGCTGCTTCAA 1410
    : : : : : : : : : : : : : : : : : : : : : : : :
257 rValAspAspSerTyrSerGlnAsnGly...GlyValIleuValAspGlu 273
    : : : : : : : : : : : : : : : : : : : : : : : :
1411 .....GCCAAAGGCGAAACCAAGGCTCGTCAGC..... 1440
    : : : : : : : : : : : : : : : : : : : : : : : :
273 snAsnSerAlaIleGlyTyrGlyAspGlyProSerSerAlaIleGlyGly 289
    : : : : : : : : : : : : : : : : : : : : : : : :
1441 .....GTGGCGGACGGTAA 1454
    : : : : : : : : : : : : : : : : : : : : : : : :
290 PheMetTyrLeuGlyLeuSerGluValThrPheAspIleAlaAspGlyL 306
    : : : : : : : : : : : : : : : : : : : : : : : :
1455 AGTCATCTTAGATCAAGGCGGACATCAAGCAAAAAACAAGCCTTTA 1504
    : : : : : : : : : : : : : : : : : : : : : : : :
306 srThrLeuValIleGlyAsnThrGlnAsnAspGlyAlaValAspSerIle 323
    : : : : : : : : : : : : : : : : : : : : : : : :
1505 GTGAATCGGCTTGTC.....ACGCGGAGGCGGACGTCGCACTGAAT 1548
    : : : : : : : : : : : : : : : : : : : : : : : :
323 laGlyThrGlyLeuIleThrLysThrGlySerGlyAspLeuValLeuAsn 339
    : : : : : : : : : : : : : : : : : : : : : : : :
1549 GCGGATTAAT..... 1557
    : : : : : : : : : : : : : : : : : : : : : : : :
340 AlaAspAsnAsnAspPheThrGlyGluMetGlnIleGluAsnGlyGluVal 356
    : : : : : : : : : : : : : : : : : : : : : : : :
1558 .....C 1558
    : : : : : : : : : : : : : : : : : : : : : : : :
356 lThrLeuGlyArgSerAsnSerLeuMetAsnValGlyAspThrHisCysG 373
    : : : : : : : : : : : : : : : : : : : : : : : :
1559 AGTTAACCAGCAAACTATTTGCGCTTCGCGCGGACGTTTGCAAT 1608
    : : : : : : : : : : : : : : : : : : : : : : : :
373 InAspAspProGlnAspCysTyr...GlyLeuThrIleGlySerIleAsp 388
    : : : : : : : : : : : : : : : : : : : : : : : :

```

```

1609 .....TTGAACGGGCAATCGCTTTCG..... 1629
    : : : : : : : : : : : : : : : : : : : : : : : :
389 GlnTyrGlnAsnGlnAlaGlnLeuAsnValGlySerThrGlnGlnThrPh 405
    : : : : : : : : : : : : : : : : : : : : : : : :
1630 .TTCCAGCCGATTTCAAAATACCGATGAAGGCGGATGATTGCAACACA 1678
    : : : : : : : : : : : : : : : : : : : : : : : :
405 eValHisAlaLeuThrGlyPheGlnAsnGlyThrLeu.....A 418
    : : : : : : : : : : : : : : : : : : : : : : : :
1679 ATCAAGACCAAAAGATCCACCGTTTACC..... 1704
    : : : : : : : : : : : : : : : : : : : : : : : :
418 snIleAspIleGlyAsnValThrValAsnGlnGlySerPheAlaGly 434
    : : : : : : : : : : : : : : : : : : : : : : : :
1705 ...ATTACGCGCAATTAAGATATTTACTACACCGGCAATTAACACAAAC.. 1749
    : : : : : : : : : : : : : : : : : : : : : : : :
435 IleIleGlnGlyAlaGlyGlnLeuThrIleAlaGlnAsnGlySerTyrVal 451
    : : : : : : : : : : : : : : : : : : : : : : : :
1750 .TTGGATAGCAAAAAAATGCTACACCGGTTGTTGGCGGAAG 1798
    : : : : : : : : : : : : : : : : : : : : : : : :
451 lleuAlaGlyAlaGlnSerMetAlaLeuThrGly.....A 463
    : : : : : : : : : : : : : : : : : : : : : : : :
1799 ATGCAACCAAAAGCAAGCGCGGCTCAATCTGAATTACCAACCGGAAGA 1848
    : : : : : : : : : : : : : : : : : : : : : : : :
463 spIleValValAspAspGlyAlaVal...LeuSerLeuGlnGlyAspAla 478
    : : : : : : : : : : : : : : : : : : : : : : : :
1849 GCGGATCCGACTTCTGCTTCCGCGGACAAAT.....TTAAACGG 1892
    : : : : : : : : : : : : : : : : : : : : : : : :
479 AlaAspLeuThrAlaLeuGlnAspAspProGlnSerIleValLeuAsnGln 495
    : : : : : : : : : : : : : : : : : : : : : : : :
1893 CAATATCACGCAACCAACGCGCAAACTGTTTTCAGCGGACGACGACAC 1942
    : : : : : : : : : : : : : : : : : : : : : : : :
495 yGlyValLeuAsnLeuSerAsp..... 502
    : : : : : : : : : : : : : : : : : : : : : : : :
1943 CGCACGCTTACAAATCATTAGAAACGGGCTGCAAAATGGAAGATATC 1992
    : : : : : : : : : : : : : : : : : : : : : : : :
503 .....PheSerThrTrpGlnSerGlyThrSerTyrAsnAspGlyLeu 516
    : : : : : : : : : : : : : : : : : : : : : : : :
1993 .....CCAAAGGAGAAATGCTGTGGGACCAACGATTGGATGCA 2030
    : : : : : : : : : : : : : : : : : : : : : : : :
517 GluValSerGlySerSerGlyThrValIleGlySerGlnAspValValAs 533
    : : : : : : : : : : : : : : : : : : : : : : : :
2031 CCGGCAATTTAAAGCGGAAATCTTCATATTCAGAGC...GGACAAACGG 2077
    : : : : : : : : : : : : : : : : : : : : : : : :
533 PheuAla...GlyGlyAspAsnLeuHisIleGlyGlyAspGlyLysAspG 549
    : : : : : : : : : : : : : : : : : : : : : : : :
2078 TGGTTTCCCGCAATGTTGCCAAAGTGCAGAGCGCATTTAAGCAAT 2127
    : : : : : : : : : : : : : : : : : : : : : : : :
549 lValIlyrValValValAspAlaSerAspGlyGlnValSerLeuAlaAsn 565
    : : : : : : : : : : : : : : : : : : : : : : : :
2128 CAGCCCAAGCAGATTTCGGTGGTGCACCGGATCAAAAGCCACAAATC.. 2175
    : : : : : : : : : : : : : : : : : : : : : : : :
566 Asn...AsnSerTyrLeuGlyThrThrGlnIleAlaSerGlyThrLeuMe 581
    : : : : : : : : : : : : : : : : : : : : : : : :
2176 .TGTAACGTTTGGAGCTGACGCGGCTGCACAAGTTTGCGAAAAACA 2224
    : : : : : : : : : : : : : : : : : : : : : : : :
581 tValSerAspAsnSerGlnLeuGlyAspThrHisTyrAsnArgGlnValI 598
    : : : : : : : : : : : : : : : : : : : : : : : :
2225 TTACCGAGATTA.....GTGATTCCTTATTGACCAAGACC 2262
    : : : : : : : : : : : : : : : : : : : : : : : :
598 lePheThrAspLysGlnGlnGlnSerValMetGluIleThrSerAspVal 614
    : : : : : : : : : : : : : : : : : : : : : : : :
2263 GACATCAGAGCAATGTCAGC.....CTGC 2288
    : : : : : : : : : : : : : : : : : : : : : : : :
615 AspThrArgSerAspAlaIleGlnHisGlyAlaArgAspIleGluMetArg 631
    : : : : : : : : : : : : : : : : : : : : : : : :
2289 CGATCAGCGTCATTTAAATCTCACAGAGACTTGCACAACCTCAACGGCAATC 2338
    : : : : : : : : : : : : : : : : : : : : : : : :
631 aaSpGlyGlnValAlaValAspAlaGlyValAspThrGlnTrpGlyAlaL 648
    : : : : : : : : : : : : : : : : : : : : : : : :
2339 TTATGTCAGCGGGA.....GACACGCAATATACGTTTACG 2373
    : : : : : : : : : : : : : : : : : : : : : : : :
648 euMetAlaAspSerSerGlyGlnHisGlnAspGlnGlySerThrLeuThr 664
    : : : : : : : : : : : : : : : : : : : : : : : :
2374 CGCAACGCGCAACCAAGCGCAACCTCAGCCTTCGCGCAATGCCAAGC 2423
    : : : : : : : : : : : : : : : : : : : : : : : :

```

```

665 LysThrGlyAlaGlyThrLeuGluLeuThrAlaSerGlyThrThrGlnSe 661
2424 AMCA.....TTTATCAAGCCACATTAAAGCGCAACACATCG..... 2460
681 rAlaValArgValGluGluGlyThrLeuLysGlyAspValAlaAspIleI 698
2461 .....GCTTCGGCAATGCTTCATTAAATCTA 2487
698 euProTyrAlaSerSerLeuTyrValGlyAspGlyAlaThrPheValThr 714
2488 .....AGCAACAAGCCGTACAAAACGCAAGCTCT 2516
715 GlyAlaAspGlnAspIleGlnSerIleAspAlaIleSerSerGlyThrI 731
2517 GAGCGTTTCGACAAACGCTAAGCAACGTAAGCCATTCCGCACTCAACG 2566
731 eaSpIleSerAspGlyThrValLeuArgLeuThr..... 742
2567 GCATATGTCCTCCGTAGCCGATTAAGCAGATTCATTTTGAACACGCCGC 2616
742 ..... 742
2617 TTTACCGGAAAAATCAGCGCGGCAAGGATAGCATTTACATTAAAGA 2666
743 .....GlyGlnAspThrSerValAlaLeuAsnAl 752
2667 CAGC.....GAATGGACGCTGCCGTGGGCAAGGAA..... 2697
752 aSerLeuPheAsnGlnLysAspGlyThrLeuValAlaThrAspGlyValT 769
2698 .....TTAGCAATTTAAC..... 2712
769 hrLeuThrGlyGluLeuAsnThrAsnLeuGluThrAspSerLeuThrTyr 785
2713 CTTCGAACAGCCACGATTACATAATCCGCTATGACACGATGGCGC 2762
786 LeuSerAsnValAlThrValAsnGlyAsnLeuThrAsnThrSerGlyAla 802
2763 AGGCGGCAAAACCGCAGTCGGGAGGACGCCGCCGCGTTCGGCGC 2812
802 lSerLeuGlnAsnGlyAlaAlaGlyAsp..... 811
2813 GTTCCCTATTATCCGTTACGCCGCCAATCTCGGACGAATCCGTTCAAC 2862
811 ..... 811
2863 ACCGTGAGGTAAACGGCAATGTGAAGCGTCAAGGAACTTCGCTTTAT 2912
812 ThrLeuThrValAsnGlyAspTyrThrGlyGlyThrLeuLeuAs 828
2913 GTTCGAACTCTTCGCG.....TACCGCAGCGGCAATTTGAAGCTGGCG 2956
828 pSerIleuLeuAsnGlyAspAspSerValSerAspGlnLeuValMetAsn 845
2957 AAAGTTCGGAAGGACATTACACTTGCGCTGCACAAAT...ACGGCG... 3000
845 lYAsnThrAlaGlyAsnThrThrValAlaValAsnSerIleThrGlyIle 861
3001 AACGAACCCGTAAGTCTCGAGCAATTGACGTAAGTGAAGAAAAAGACAA 3050
862 GlyGluProThrSerThrGlyIleLysValAlaAspPheAlaAlaAspPr 878
3051 CACACCGGTTCGCAAAATCTTAATTTCAACCTGCAAAAGCA...CAGC 3097
878 oThrGlnPheGlnAsnAlaGlnPheSerLeuAlaGlySerGlyTyrV 895
3098 TCGATGCGCGGCGCATGCGCTTATCAGCTTATCCGCAAGACGGCAGTTC 3147
895 aLAsnMetGlyAlaTyrAspTyrThrLeuValGluAspAsnAsnAspThr 911
3148 CGCGTCGATTAATCCGTCGAAGAACAGAGGCTTCCGACAAACCTGGCAA 3197
|||
912 TyrLeu..... 913
3198 GCGGAGAGAAACAGAGCGCGCTTACGGCAAAACAGGCAACACTTCGCC 3247
913 ..... 913
3248 CCAACACAGCGGAAAAAGACAAACGCGCAAAAGCCTTAGCGCGTGAAT 3297
913 ..... 913
3298 GCGCGCGGCGCAATGCCACCGAAAGCAGCAAAAGTGTGCCGAACCGCG 3347
914 .....ArgSerGlnGluValThrProProSe 922
3348 CCGGACGCGAGCGGGGAAAATGCCGATTAATGACGCGGAGAGAGAGA 3397
922 r..... 922
3398 AAAACGGGTGACGCGGATTAAGACACCGCCTTGCGAAGACGCGGAA 3447
923 .....ProProAspProAsp..... 927
3448 GCGGAACCGCGCGCTACCAACGCGCTTCCCGCGCGCGCGCGCGCG 3497
928 .....ProThrProAsp..... 931
3498 CCGGATTTGCGGCAACCGCACCGCCCAACCGCAACCGCACGCGCG 3547
932 .....ProAspProThrProAspProAspProThrProAspProG 945
3548 ACCTGATCAGCGGTTATGCCAATAGCGGTTTGAGTGAATTTCCGCGCAG 3597
945 lProThrProAlaTyrGlnProValLeuAsnAlaLysValGlyGlyTyr 961
3598 CTCACAGCGTTTTCGCGCTACAGGACGAATTCGACGCGCTGTCCGA 3647
962 LeuAsnAsnLeuArgAlaAla.....AsnGlnAlaPheMetIle 974
3648 AGACGCGCGCAACGCGCTTGGCAAGCGCG..... 3678
974 tGluArgArgAsnIleAlaGlyLysAspGlyGlnThrLeuAsnLeuArgV 991
3679 ..ATCGGAGACACCAACACTACCGTTGCGAAGATTTCGCGCGCTACGCG 3726
991 allIleGlyLysPyrThrIleThrAlaAlaGlyLeuAla..... 1005
3727 CAACAAACGCACTGCGCCCAATCGGTATGCAGAAAAACCTCGGACGCG 3776
1006 GlnHisGluAspThrSerThrValGlnLeuSerGlyAspLeuPheSerG 1022
3777 GCGC.....GTTCGCAATCCG.....TTT 3796
1022 yArgTyrGlyThrAspGlyLysTrpMetLeuGlyIleValGlyGlyTyrS 1039
3797 CGCACAACCGGACGGAACACCTTCGACGACGCGATCGCAACTCGCGCA 3846
1039 erAspAsnGln...GlyAspSerArgSerAsnMetThrGlyThrArgAl 1054
3847 ...CGGCTCCCAACGCGGCTTCGCG.....CA 3875
1055 AspAsnGlnAsnHisGlyTyrAlaValAlaGlyLeuThrSerSerTrpPheG 1071
3876 ATACGGATGGGACGATTCGACATCGGCATCGACGCGGCGCGCGGTGTA 3925
1071 nHisGlyAsnGlnLysGlnGlyAlaTrpLeuAspSerTrpLeuGlnTyrA 1088
3926 GT.....ACGCGCAGCGCTTCAGACGCGCATCGAGCAAAATCCGCGC 3969
1088 lAtThrPheSerAsnAspValSerGluGlnGluAspGly..... 1100
3970 CCGGTGCTGCATTAC.....GGCATTCAGGCAAGATACCGCGCAG 4010
1101 ...ThrAspHisTyrHisSerSerGlyIleIleAlaSerLeuGluAlaG 1116
```



```
1831 .AATTACCAACCGGAGAGGACGACATTACTGCTTTCCGGCGGAA 1879
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
762 nLysLysGlnThrAspAsnThrSerPheThr.....IleSerGlyGlyP 777
1880 CAATTAAACGCAATATACGCAACCAACGCAACGCAATGTTTTCACG 1929
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
777 heSerTyThrGlyGlyValAspLysValGlySerLysAlaAspPheGln 793
1930 GCGAGACCGACACCGCGCTACATCATTTAGGAGCGGGTGGTCAAA 1979
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
794 .....TyrAspLysGlnHisThrGlnThrGluValThrLys 805
1980 AATGAGAGTATCCACAGAGAAATCGTGGAC..... 2016
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
805 sAsnArgGly...SerGlnThrGluValAlaGlyAspLeuThrIlePheTha 821
2017 ..AACGATTGGATCGACCGCACATTAAAGCGGAAATTCATATTCAG 2064
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
821 lAsnLys.....AspLeuLeuHisGlyLysAlaSerHisHisValGlu 835
2065 GCGGACAGACGGTGGTTCGCCGCAATGTTCGCAAGTGGAGGGGATTG 2114
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
836 GlYArgTyThrGlnGlnSerGlyGlnLysIleGln..... 846
2115 GCATTTAAGCAATACAGCCCAAGACATTTCGGTGTGCGACCGCATCAA 2164
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
847 .HisLeu.....AlaValAsnAspS 853
2165 GCGACACATCTGTACACGTTG..... 2187
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
853 ergLInThrSerLysThrAspSerLeuAsnValGlyIleAspValGlyVal 869
2188 .....GACTGACGGGTCTGCAAGTTGTACGCAAAACCATACCGA 2231
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
870 AsnLeuAspTyThrSerGlyValThrLysProValLysLysAlaIleGlnAs 886
2232 CGAT...AAGTATGCTTCATTGACAGACGACATCAGAGGCAATG 2278
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
886 pGlyValAsnThrThrLysProGlyAsnAsnThrAspLeuThrLysLysV 903
2279 TCAGCCTTCCGATCAC...GCTCATTTA..AATTCACAGGATTTGCC 2322
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
903 alThrAlaArgAspAlaIleAlaAsnLeuAlaAsnLeuSerAsnLeuGlu 919
2323 ACACTCACGCGCAATCTTAGTCAGGC..... 2349
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
920 ThrProAsnValGlyValGluValGlyIleLysGlyGlySerGlnGly 936
2349 ..... 2349
936 nSerGlnThrAspSerGlnAlaValSerThrSerIleAsnAlaGlyLysI 953
2350 .....GAGACACGACATACG 2367
953 leAspIleAspSerAsnAsnLysLeuHisAspGlnGlyThrHisIleTy 968
2368 GTTACGCGCAACGCCCAACGCAACCTCAGCTCGTGGCAATGC 2417
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
969 .....GlnSerThrGlnGlnGly...IleSerLeuThrAlaAsnTh 981
2418 C.....CAAGCAATTTAATCAG 2437
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
981 ThrHisThrSerGlnAlaThrLeuAspLysHisGlnThrThrPheHisGln 998
2438 CCACATTAAACGCAACATCGCTTCGACAAATGCTTCATTAAATGTA 2487
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
998 hrLysGlyGlyGlnIleGlyValSerThrLysThrGlySerAspIle 1014
2488 AGCAACAAACGCGGTACAAACGCGAGTGTGACGTTTCCGCAACGCT. 2535
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1015 ThrValAlaIleLysGlyGlyGlnThrThr.....AspAsnAlaLe 1029
2536 .....AAGCAAAACGTAAAGCAATTCGCGACTACAGCGCAATGTCT 2575
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1029 uMetGluThrLysAlaLysGlySerGlnPheThrSerAsnLysLysIleS 1046
2576 CCCTAGCCGATTAAGCACTATTCATTGTAAGAAACAGCCGCTT. 2619
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1046 erLleAsnValGlyGluAsnAlaHisTyThrGlyGlnGlyAlaGlnPheAspAla 1062
2620 ..ACGGAATAA.....ATCAGCGCGCGCAAGGATACGGCATTAACATT 2660
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1063 GlnLysGlyLysThrValIleAsnAlaGly..... 1072
2661 AAAGACACGCAATGACGCTGCCGTGGCGACGGAATTAGCAATTTAA 2710
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1073 .....GlyAspLeuThr 1076
2711 ACCTTGACACCGCCACCATTAACCTCAATTCGCTATACACAGATGCG 2760
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1076 hrLeuAlaGlnAlaThr..... 1081
2761 GCAGGCGCGCAACCGGACGATGCGGACAGATGCGCGCGCGCTTCGG 2810
1081 ..... 1081
2811 CGGTTCCCTATTATCCGTTACGCGCGCACTGGCGCAAAATCCGTTCA 2860
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1082 .....AspThrHisSerGlnSerGlnSer 1090
2861 ACACGCTGACGGTAACGCGCAATTTGAACGGTCAAGGACATTCGCTTT 2910
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1090 sn.....ValAsnGlySerAlaAsn..... 1096
2911 ATGTGCGACACTTCGCGCTACGCGACGCGCAATTTGAAGTGGCGGAA 2960
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1097 .....LeuLysValGlyThrThr 1102
2961 TTCGAGGACACTTACACCTTGCGCTGCACAATATTCGCGCAAGAACCC 3010
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1102 rProGlnSerLysAspTyThrGlyGlyPheAsnAlaGlyThrThrHisH 1119
3011 TAACTTCGCAATTAAGC.....GTAGTGAAGGAAAGAAC 3048
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1119 lSerLysGlnGlnThrThrAlaLysValGlyThrIleThrGlySerGln 1135
3049 AACACACCGCTGTCGCAAAATCTTAATTCACCCCTGCAAAACACACGT 3098
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1136 GlyIleGluLeuAsnAlaGlyHisAsnLeuThrLeuGlnGlyThrHisLe 1152
3099 CGATGCCGCGCATGGCGTTATGAGTTATCCGCAAGACGCGAGTTCC 3148
1152 u..... 1152
3149 GCGTGCATTAATCCGGTCAAGAACACAGGCTTTCGCAAACTCGGCAAG 3198
1152 ..... 1152
3199 GCGGAGAGAAACAGAGCGGCTTGACGCGCAAAA.....CAGGACAA 3239
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1153 SerSerGluGlnAspIleAlaLeuAsnAlaThrAsnLysValAspLeuG 1169
3240 ACTTGCGCGCAACACAGCGGAGGAAAAAGACAGCGCAAGGCTTGACG 3289
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1169 nSerAlaSerSerGlnHisThrGlnLysGlyAsnAsnLeuSerGlyGly 1186
3290 CGCTGATTGGGCGCGGCGCAATGCGCAAGGAAAGCAGAAATGTT... 3336
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1186 alGlnAlaGlyPheGlyLysLysMetThrAspAlaSerValAsn 1202
3337 .....GCGGACCGGCG 3347
1203 GlyLeuGlySerAlaGlnPheAlaIleGlyLysGlnAspGluLysSerVal 1219
3348 CCGGACAGGACGCGGG.....GAAAATGCCGCGCATTTATGACGCGGAG 3391
```

[illegible]

```

seq_documentation_block:
ID      PERT_BORPA      STANDARD;      PERT;      922 AA.
AC      P24328;
DT      01-MAR-1992 (Rel. 21, Created)
DT      01-MAR-1992 (Rel. 21, Last sequence update)
DT      01-OCT-1996 (Rel. 34, Last annotation update)
DE      Pertactin precursor (Outer membrane protein P.70) (P.95) .
GN      PRN.
OS      Bordetella parapertussis.
OC      Bacteria; Proteobacteria; beta subdivision; Alcaligenaceae;
OC      Bordetella.
OC      NCBI_TaxID=519;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=ISOLATE CN2591;
RX      MEDLINE=91251771; PubMed=2041476;
RA      Li L.J., Dougan G., Novotny P., Charles I.G.;
RT      "P.70 pertactin, an outer-membrane protein from Bordetella
RT      parapertussis: cloning, nucleotide sequence and surface expression in
RT      Escherichia coli."
RL      Mol. Microbiol. 5:409-417(1991) .
CC      -1- FUNCTION: AGGLUTININ THAT BINDS TO EUKARYOTIC CELLS; A PROCESS
CC      MEDIATED BY THE R-G-D SEQUENCE. PERTACTIN MAY HAVE A ROLE IN
CC      BACTERIAL ADHESION, AND THUS PLAY A ROLE IN VIRULENCE.
CC      -1- SUBUNIT: MONOMER.
CC      -1- SUBCELLULAR LOCATION: Outer membrane.
CC      -1- DISEASE: MAY CONTRIBUTE TO THE DISEASE STATE OF WHOOPING COUGH.
CC      -1- MISCELLANEOUS: SYNTHESIZED ONLY IN THE PRESENCE OF LOW MG(2+)
CC      CONCENTRATIONS.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; X54547; CA38419.1; -
DR      EMBL; A26124; CA01786.1; -
DR      EMBL; A19182; CA01454.1; -
DR      PIR; S15204; S15204.
DR      PIR; S14659; S14659.
DR      InterPro: IPR003992; pertactin.
DR      InterPro: IPR003991; pertactin_related.
DR      PRINTS; PR01482; PERTACTIN.
DR      PRINTS; PR01484; PERTACTINFAMILY.
KW      Outer membrane; Signal; Virulence; Repeat.
FT      FT SIGNAL 1 34 POTENTIAL.
FT      CHAIN 35 922 PERTACTIN (P.70) .
FT      CHAIN 35 647 POTENTIAL.
FT      PROPEP 648 922 CELL ATTACHMENT SITE (INVOLVED IN
FT      SITE 260 262 ADHESION TO VARIOUS EUKARYOTIC CELL
FT      LINES) .
FT      DOMAIN 266 290 4 X 5 AA TANDEM REPEATS OF G-G-A-V-P.
FT      REPEAT 266 270 1.
FT      REPEAT 271 275 2.
FT      REPEAT 276 280 3.
FT      REPEAT 281 285 4 (APPROXIMATE) .
FT      DOMAIN 575 603 9 X 3 AA APPROXIMATE REPEATS OF P-Q-P.
SQ      SEQUENCE 922 AA; 95178 MW; 3DF7BC58D4712478 CRC64;

alignment_scores:
Quality: 275.50 Length: 1162
Ratio: 0.566 Gaps: 34
Percent Similarity: 41.910 Percent Identity: 21.687

alignment_block:
US-09-303-518D-653 x PERT_BORPA ..
Align seg 1/1 to: PERT_BORPA from: 1 to: 922

```



```

481 ArgMetLanValPheAlaSplLeuGlyLeuSerSplLysLeuValAlaIle 497
|||||
2955 GGAAAGTTCCGAGAGGCACTTACCTGGCTGGTCAACAAATACGGCAACG 3004
|||||
497 LArgSplAlaSerGlyGlnHisArgLeuTrpValLArgSplSerGlySerG 514
|||||
3005 AACCCGTAAGTCTGAGCAATTGACGGTATGGAAGAAAGAACACACCA 3054
|||||
514 LArgSplAlaSerGlyGlnHisArgLeuTrpValLArgSplSerGlySerG 526
|||||
3055 CCGCTGTCCGAAATCTTAATTCACCTGCACAAACGAA.....CAGCT 3098
|||||
527 ProAlaGlySerAlaAlaThrPheThrLeuAlaHisnLysAspGlyLysVa 543
|||||
3099 CGATGCGCGCGCAGGCGTTATCAGCTTATCCGCAA...GACGGCGAGT 3145
|||||
543 LAspLLeuGlyThrTyrArgTyrArgLeuAlaAlaHisnLysnGlyGlnT 560
|||||
3146 TCCTGGCTGCATTAATCCGGTCAAGAACAAAGCTTCCGACAACTCGGC 3195
|||||
560 TrpSerLeu.Val..... 563
|||||
3196 AAGCGGGAGAAACAGAGCGCGCTTACGCGCAAAACAGGCAACACTTGC 3245
|||||
563 ..... 563
|||||
3246 CGCCAAACACAGCGCGGAAAGAACAGCCCAAGCCTTGACGGCGCTGA 3295
|||||
563 ..... 563
|||||
3296 TTGCGCGCGGCGCAATGCAACGAAAGGCAAGAAAGTGTCCGCAACG 3345
|||||
564 .....GlyAlaLysAlaLysProAlaProLysProAlaPro..GI 576
|||||
3346 GCGCGGAGGAGGCGCGGAAATCCGCGCTTATGCAAGCGGAGGAGAA 3395
|||||
576 nProGlyProGlnProGlyProGlnProPro..... 586
|||||
3396 GAAAAAAGCGGTGCAAGCGGATTAAGACACCGCTTGGGAAACAGCGCG 3445
|||||
586 ..... 586
|||||
3446 AAGCGGAAACCGCGCGGCTACACCGCTTCCCGCGCGCGCGCGCGCC 3495
|||||
587 .....GlnProProGln..ProProGlnProProGln 597
|||||
3496 CGCGCGGATTTGCGCAACCGGAGCGCCAA.....CGGCAACCGCA 3536
|||||
597 rProPro.GlnProProGlnArgGlnProGlnAlaProAlaLysProLys 613
|||||
3537 ACCGCGAGCGGACCTGATCAGCGCTTATGCAATAGCGGTTTGAGTGAAT 3586
|||||
613 oAlaGlyArgGlyLeu..SerAlaAlaAlaHisnAlaAlaValaLanThrG 629
|||||
3587 TTTCCGCGCAGCTCAACAGCGTT.....TTGCGCGTACAGGACGAATTG 3630
|||||
629 LysLysAlaLysLeuAlaSerThrLeuTrpTyrAlaLysSerHisnLysLeu 645
|||||
3631 GACCGCGGTGTTCGCAAGACCGCGCGCAAC.....GCGGTTCG 3668
|||||
646 SerLysArgLysGlyLysLeuArgLysLeuHisnProAlaAlaGlyGlyAlaTr 662
|||||
3669 GACAAACCGGCTACCGGAGCACCAACACTACGCTTCGCAAGATTTCCGCG 3718
|||||
662 pGlyArgGlyPhe.....A 667
|||||
3719 CCTACCGGCAACAAACCGACCTGCGGCAATAGGTATGCAAAAAACCTC 3768
|||||
667 LArgLysArgGlyGlnLysLeuArgLysLeuHisnProAlaAlaGlyGlyAlaTr 673
|||||
3769 GGCAGCGGCGCGGTGCGCATCTGTTTTCGCAACCGGAGCGGAAACAC 3818
|||||

```

```

674 .....AsnArgAlaGlyArgTr 679
3819 CTTCGAGCAC.....GGCAGCGGCAACTGGGCGAC 3847
|||||
679 gPheAspArgLysLysValAlaGlyPheGlyLeuGlyAlaAspHisAlaVala 696
|||||
3848 GCGTCCCGCCGAGGTGCC.....GTTTCGGGCAATACGCAATCGGACG 3891
|||||
696 LysAlaLysGlyArgGlyTrpHisLeuGlyGlyLeuLysGlyTyrThrArg 712
|||||
3892 TTGACATCGCGCATCAGCGCGGCGCGGCTTTAGTACGGGCAAGCTT... 3939
|||||
713 GlyAspArgGlyPheThrGlyAspGlyGlyGlyHisnHisAspSerValHis 729
|||||
3939 ..... 3939
|||||
729 sValGlyGlyTyrAlaThrTyrThrLysAlaHisnSerGlyPheTyrLeuAspA 746
|||||
746 LThrLeuArgAlaSerArgLeuGlyLysAspPheLysValAlaGlySer 762
|||||
3943 GACGCG.....ATCAGAGGCAAAATCCGCGCGCGCTGCGATTAACG 3986
|||||
763 AspGlyTyrAlaValLysGlyLysTyrArgThrHisGlyValGlyValSe 779
|||||
3987 CATTCAGGCAAGATACCGC...GCAGGTTTCGGCGGATTCGGCATCGAAC 4033
|||||
779 LLeuGlnAlaGlyArgArgPheAlaHisnAlaAspGlyTrpPheLeuGln 796
|||||
4034 CGCAGCATCGCGGCAACCGCGCTATTTGTCGCAAAAGCGATTCGCGATAC 4083
|||||
796 rGlnAlaGlyLeuAlaValAlaPheArgValGlyLysLysAlaTyrAlaGlyAla 812
|||||
4084 GAAAC...GTCAATATCGCCACCGCGCGCTGTCATTCAGCGCTACG 4130
|||||
813 AlaHisnLysLeuArgValArgAspGlyGlyLysSerValLeuGlyTr 829
|||||
4131 CGCGGCGATTAAGGCGATTAATTCATTCAGCGCGGCAACACTTCCA 4180
|||||
829 LLeuGlyLeuGlyValGlyLysArgGlyLeuLysAlaGlyLysArgGln 846
|||||
4181 TCAGCGCTTATTTGAGCGCTGTCGATTCAGGATTCGCGCAAGTGC 4230
|||||
846 ALGlnProTyrLysLysAlaSerValLeuGlnLysPheAspGlyAlaGly 862
|||||
863 ThrValArgThrHisnGlyLysLeuAlaHisnArgThrGlyLeuArgGlyThrArg 879
|||||
4281 CAGTGGCGAATGGGCGGTAAACCGCGCAATCAAGTTTCAGCGCTGCC 4330
|||||
879 G...AlaGlyLeuGly.....LeuGly 886
|||||
4331 TTCAGCGTTCGCGCGCGGCAAGGCGCGGCAATTCGCAAGC.....CAGCAC 4374
|||||
886 eAlaAlaAlaLysLeuGlyArgGlyHisSerLeuTyrAlaAspTyrGlyLys 902
|||||
4375 AGCGGCGGCAATCAATTAAGCTACCGCTGG 4404
|||||
903 SerLysGlyProLysLeuAlaMetProTyr 912
|||||
seq_name: SwissProt_40:PERT_BORBR
seq_documentation_block:
ID PERT_BORBR STANDARD; PRT; 911 AA.
AC G03035;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Pertactin precursor (Outer membrane protein P.68) (P.94).
GN PRN.
OS Bordetella bronchiseptica (Alcaligenes bronchiseptica).

```

```

OC Bacteria; Proteobacteria; beta subdivision; Alcaligenaceae;
OC Bordetella.
OX NCBI_TaxID=518;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 35-44.
RC STRAIN-ISOLATE CN7531;
RA MEDLINE=92407514; PubMed=1527510;
RA Li J.L., Fairweather N.F., Novotny P., Dougan G., Charles I.G.;
RT "Cloning, nucleotide sequence and heterologous expression of the
RT protective outer-membrane protein P.68 pertactin from Bordetella
RT bronchiseptica";
RL J. Gen. Microbiol. 138:1697-1705(1992).
CC -1- FUNCTION: AGGLUTININ THAT BINDS TO EUKARYOTIC CELLS; A PROCESS
CC MEDIATED BY THE R-G-D SEQUENCE. PERTACTIN MAY HAVE A ROLE IN
CC BACTERIAL ADHESION, AND THUS PLAY A ROLE IN VIRULENCE.
CC -1- SUBUNIT: MONOMER.
CC -1- SUBCELLULAR LOCATION: Outer membrane.
CC -1- DISEASE: MAY CONTRIBUTE TO THE DISEASE STATE OF WHOOPING COUGH.
CC -1- MISCELLANEOUS: SYNTHESIZED ONLY IN THE PRESENCE OF LOW MG(2+)
CC CONCENTRATIONS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X54815; CAA38584.1; -
DR EMBL; A19180; CAA01453.1; -
DR PIR; A47675; A47675.
DR InterPro; IPR003992; pertactin.
DR InterPro; IPR003991; pertactin_related.
DR PRINTS; PR01482; PERTACTIN.
DR PRINTS; PR01484; PRACTINFMALY.
KW Outer membrane; Signal; Virulence; Repeat.
FT SIGNAL 1 34
FT CHAIN 35 911 P.94.
FT CHAIN 35 712 PERTACTIN (P.68).
FT PROPEP 713 911 POTENTIAL.
FT SITE 260 262 CELL ATTACHMENT SITE (INVOLVED IN
FT ADHESION TO VARIOUS EUKARYOTIC CELL
FT LINES).
FT SITE 701 703 CELL ATTACHMENT SITE (POTENTIAL).
FT DOMAIN 266 275 3 X 5 AA TANDEM REPEATS OF G-G-A-V-P.
FT REPEAT 266 270 1.
FT REPEAT 271 275 2.
FT REPEAT 276 280 3 (APPROXIMATE).
FT REPEAT 280 280 7 X 3 AA REPEATS OF P-O-P.
FT DOMAIN 570 601 307BDF6EC2D987A1 CRC64;
SO SEQUENCE 911 AA; 93995 MW; 307BDF6EC2D987A1 CRC64;

alignment_scores:
Quality: 274.00 Length: 1149
Ratio: 0.584 Gaps: 50
Percent Similarity: 40.818 Percent Identity: 20.540

alignment_block:
US-09-303-518d-653 x PERT_BORBR ..

Align seg 1/1 to: PERT_BORBR from: 1 to: 911

1306 GGCGGGGGCTCATATGATGATGACGATACGTTACTTGGGAAGTAAA 1355
|||||
58 G1YAlaG1YValArgThrAlaThrGlyThrThrIle.....LysValSe 72
1356 CGGCGTGGCAAGAGACGGCTGTCCAAAATCGCAAGAGACGCGTGTGG 1405
|||||
72 rGly.....ArgGlnAlaGlnGlyValLeuLeuG 82
1406 TTCAGCCAAAGGGGAAAAACCAAGGCTGGTACGCTGGCGGACGCTAAA 1455
::: |||||

```

```

82 LuAsnProAlaIaGlu.....LeuArgpHeGlnAsnGlySer 94
1456 GTCAATCTTAGATCAAGCGCGAGATCAAGCGCAAAAACAGCCTTAG 1505
|||||
95 ValThrSerSerGlyGlnLeuPheAspGlnGlyValArgArgPheLeu.. 110
1506 TGAATCGGCTTGGTCAAGCGGAGGAGGAGGTCGATCAATACCCGATA 1555
|||||
111 .....GlyThrValThrValLysAla.... 117
1556 ATCAGTTCACCCCGACAAACTGATTTCGGCTTTCGGCGGACGTTTG 1605
|||
118 .....GlyLysLeu 120
1606 GATTGAAGGCGCATTCGCTTTCGTTCCACCGCATCAAAATACCGTGA 1655
|||
121 ValAlaAspHisAlaThrLeuAlaAsnValSerAspThrArgAspAspAs 137
1656 AGGGCGGATGATGTCAACCAACCAATCAAGACAAAGAAATCCACGCTT... 1701
|||
137 pGlyIleLeuLeuThrValAlaIaGlyGlnGlnAlaIaSerIleAla 154
1702 .....ACCATTTACGCAATTAAGATATTTACTACACCGGCAATACAC 1746
|||||
154 spSerThrLeuGlnGlyAlaGlyGlyValArgValGluArgGlyAlaAsn 170
1747 AACTTGATACCAAAAAAGAAATTCCTACACGCTTGTTGGCGGAGAA 1796
|||
171 ValThrValGlnArg..... 175
1797 AGATGCAACCAAAAGAACGGCGGCTCAATCTGAAT.....T 1834
|||
176 .....SerThrIleValAspGlyGlyLeuHisIleGlyThrLeuGlnPro 191
1835 ACCAACCGGAGAAAGAGGATCGCACTTACTGCTTCGCGGCAACAAT 1884
|||||
191 eugInProGluAspLeuProSerArgValValLeuGlyAspThrSer 207
1885 TTAAACGGCAATATCAACGCAAAACCAAGCAACCTGTTTTCAGCGGAG 1934
|||
208 ValThrAla.....ValProAlaSerGly..... 215
1935 ACCGACACCGCACGCTTACATCAATCTTAGAAGCGGTGTCAAAATGG 1984
|||
216 .....AlaProAlaIaValSerValPheGlyAlaAsnGlnLeuThrVal 231
1985 AAGGTATCCCAAGAGAAATCGTGGGACAAACGATGATCGACCGC 2034
|||
231 spGly..... 232
2035 ACATTTAAAGCGGAAACTTCATATTCAGGCGGCAACCGGTGTTTC 2084
|||||
233 .....GlyHisIleThrGlyGlyValArgAlaAla.... 241
2085 CCGCAATGTTGCCAAAGTGGAGAGCGATTTGACAAATACAGCGCC 2134
|||
242 .....GlyValAlaIaIaMetAspGly..... 248
2135 AAGCATTTTCGGTGTGCGACCGCATCAAGACACACAAATCTGTACAGT 2184
|||
248 ..... 248
2185 TCGAGCTGAGAGGGTGTGACAACTTTGACGAAAAAACATTACCGGCA 2234
|||
248 ..... 248
2235 TAAAGTATGCTTCATTGAGCAAGACGACATCAGAGGCAATGTGAGCC 2284
|||
249 .....AlaIleValHisLeuGlnArgAlaThrIleArg..... 259
2285 TTGCCGATCAACGCTCATTTAAATCTCACAGGACTTGGCCACACTCAACGGC 2334
|||
260 .....ArgGly 261

```

```

2335 AACTCTAGTCAGCGGAGACAGCACTATACGTTACGCGCAAC 2384
      ::::: ||||| :::::
262 AspAlaProIaIaGlyLy.....AlaValProGlyGlyAlaVa 274
2385 CCAAAACGGC.....A 2395
      :::::
274 LProGlyGlyPheGlyProLeuLeuAspLyrPtryGlyValAspVal 291
2396 ACCTGACCTCGTGGCAATGCCCAAGCAACATTTAATCAACGCCATTA 2445
      :: ||| ||| ||||| ::::
291 eAspSerThrValAspLeuAlaGlnSerIleValIGluAlaProGlnLeu 307
2446 AACGGCAACACATCGGCTTCGGACAATGCTTCATTTAATCTAAGC.... 2490
      ::::: ||||| :::::
308 GlyAlaAlaIleArgAlaGlyArgGlyAlaArgValThrValSerGlyGI 324
2491 .....ACACAGCGCTACAAAACGGC..... 2511
      :::::
324 ySerLeuSerAlaProHisGlyAsnValIleGluThrGlyGlyAlaAla 341
2512 .....AGTCTGACGCTTCGGAACAAC 2532
      ||||| :::::
341 rGArgPheProProAlaSerProLeuSerIleThrLeuGlnAlaGly 357
2533 GCTAAGCAAC.....GT 2546
      ||||| :::::
358 AlaArgAlaGlnGlyArgAlaLeuLeuTyrArgValLeuProGluProVa 374
2547 AAGCCATTCCGACATCAAGCAATGTCTCCCTACCGCGTAAGCAAGTAT 2596
      ||||| :::::
374 LysLeuThrLeuAlaGlyGlyAlaGlnGlyGlnGlyAsp.....IleV 389
2597 TCCATTTGAAAAACAGCCCTTACCGGAAATACAGCGCGGCAAGAT 2646
      ||||| :::::
389 alaIaThrGluLeuProProIleProGlyAlaSerSerGlyProLeuAsp 405
2647 ACGGCATTACCTTAAGACAGCAAGTACAGCTGCGCGGCGACGCA 2696
      :::::
406 ValAlaIleu...AlaSerGlnAlaIaArgTrpThr...GlyAlaThrArgAl 420
2697 ATTGAGCAATTAACCTTGACACAGCCACCATATACACTCATTCGCCGT 2746
      :::::
420 aValAspSerLeuSerIleAspAsnAlaThrTrpValMetTrpAsp... 435
2747 ATCGACACGATCGGCGAGCGGCAAAACCGCAGTGGCGGAGATCGCGC 2796
436 .....AsnSerAsnValGlyAlaLeu 442
2797 CGCGCGCGTTGCGCGCTCCCTATTATCCGTTACGCGCGCACTTCGCG 2846
      ||||| :::::
443 ArgLeuAlaSerAspLysr...ValAspPheGlnGlnProAlaGluAl 458
2847 AGAATCCGCTTCAACAGCGTACGCTTAACGCAATTTGAAGCGTCAGG 2896
      :::::
458 a...GlyArgPheLysCysLeuMetValAsp...ThrLeuAlaGlySerG 473
2897 GAACATTCGCTTTATGTGGAACCTTCGCGCTACCGGCAAGCGCAATTTG 2946
      ||||| :::::
473 LysLeuPheArgMetAsnValAlaPheAlaAspLeuGlyLeuSerAspLysLeu 489
2947 AAGCTGCGGAAAGTTCCGAAGGCACTTACACCTTGCTGTCAACAATAC 2996
      :::::
490 ValValMetArgAspAlaSerGlyGlnHisArgLeuLeuValArgAsnSe 506
2997 CGGCAACGAAACCCGTAAGTCTGAGCAATGTGACGTAAGTGAAGAAAG 3046
      :::::
506 rGlySerGluProAlaSerGlyAsnThrMetLeuLeuValGln..... 520
3047 ACAACACACCGCTGTCGCAAAATCTTAATTTACCTCGCAAAACGAA... 3093
      ||||| :::::
521 .....ThrProArgGlySerAlaIaIaThrPheThrLeuAlaAsnLysAsp 535

```

```

3094 ...CACGTGATGCCGCGCATGGCGTATTATACGTTATCCGCAA...GA 3137
      ||||| :::::
536 GlyLysValAspIleGlyThrTyrArgTyrArgLeuAlaIaAsnGlyAs 552
3138 CGGCGAGTTCGCGCTGCATATCCGGTCAAGAAGAACAGACTTTCGACGA 3187
      :::::
552 nGlyGlnTrpSerLeu...Val..... 558
3188 AACTCGGCAAGCGCGGAGAAACAGAGGCGCGCTTGACGCAAAACAGCA 3237
558 ..... 558
3238 CAACCTCCGCGCAACACAGCGGAGAAACAGACGCGCAAGACCTTGA 3287
558 ..... 558
3288 CGGCGTATTTGGCGCGCGGCAATGCCACCGAAAGCAAGATGTTG 3337
559 .....GlyAlaLysAlaProProAlaProLysProAla 569
3338 CCGAACCAGCGCGGAGGCGGAGGAAATGCGGCAATATGACAGCG 3387
      ||||| :::::
570 Pro...GlnProGlyProGln..... 575
3388 GAGGAAGAGAAAAACGGGTGACAGCGGATTAAGACACCGCTTGGCGAA 3437
575 ..... 575
3438 ACAGCGGAGAACGGAACCGCGCGGCTACACCGCTTCGCGGCGCC 3487
576 .....ProGlyProGlnProProGln...ProProGlnP 586
3488 GCCGCGCGCGCGGATTTCCGCAACCGGAGCCCAACCGCAACCCAA 3537
      ||||| :::::
586 oProGlnProProGlnArgGlnProGluAla...ProAlaProGlnProPro 602
3538 CCGCAGCGGACCTGATCAGCCGTTATGCCAATTAAGCGTTGAGTGAATT 3587
      ||||| :::::
603 AlaGlyArgGluLeu...SerAlaAlaIaAsnAlaValaLysnThrGI 618
3588 TTCGCGCACGCGTCAACAGCGTT.....TTCGCGGTACAGCAAGAAATTGG 3631
      :::::
618 yGlyValGlyLeuAlaSerThrLeuTrpTyrAlaIaGlnSerAsnAlaLeuS 635
3632 ACCGCGTGTTCGCGAAGACCGCGCAAC.....GCCGTTTGG 3669
      :::::
635 eLysArgLeuGlyGluLeuArgLeuAsnProAspAlaGlyGlyAlaTrp 651
3670 ACAAGCGGATCGGCGACACCAACACTACCGTTCCGAATTTCCGCGC 3719
      :::::
652 GlyArgGlyPhe.....Al 656
3720 CTACCGCAACAAACCGACCTGGCGCAATTCGATGACAGAAACCTCG 3769
      ||||| :::::
656 aGlnArgGlnGlnLeuAsp..... 662
3770 GCAGCGGCGCGTTCGCGCATCTGTTTTCGCAACACCGGAGCAACACC 3819
      ||||| :::::
663 .....AsnArgAlaGlyArgArg 668
3820 TTCGACGAC.....GCGATCGGCACTCGGACAG 3848
      ||||| :::::
669 PheAspGlnValAlaIaGlyPheGluLeuGlyAlaAspHisAlaValAl 685
3849 GCTTGCCACAGGTGCC.....GTTTCGGGCAATACGGCATGGCAAGT 3892
      :::::
685 aValAlaGlyGlyArgTrpHisLeuGlyGlyLeuAlaGlyTyrThrArgG 702
3893 TCGCATTCGCGATCAGCGCGCGCGGCTTTAGTACGCGCAAGCTT... 3939
      ||||| :::::
702 LysAspArgGlyPheThrGlyAspGlyGlyGlyHisThrAspSerValHis 718
3939 ..... 3939

```



```

719 ValGlyGlyTyrAlaThrTyrIleAlaAsnSerGlyPheTyrLeuAspAl 735
3940 .....TCAG 3943
735 aThrLeuArgAlaSerArgLeuGluAsnAspPheLysValAlaGlySerA 752
3944 ACGGC.....ATCAGAGCGCAAAATCCCGCGCGCTGTCGATACGAGC 3987
||||| .....||||| |||
752 spGlyTyrAlaValValGlyGlySerTyrArgThr.....HisGly 764
3988 ATTAGCGACAGATACCGCGCAGGT.....TTCGGCGGATT 4022
||||| .....|||||
765 ValGlyAlaSerLeuGluAlaGlyArgArgPheAlaHisAlaAspGlyTr 781
4023 CGGATGCAACCGCAGCATCGCGCGCAACGCCCTATTGTCGCAAAACCGC 4072
: .....: |||
781 PheLeuGluProGlnAlaGluLeuAlaValPheArgValGlyGly 798
4073 ATTACCGATACGAAAC...GTCAATATCGCCACCGCGCGCTTGCTTC 4119
: .....: |||
798 eTyrArgAlaAlaAsnGlyLeuArgValArgAspGlyGlySer 814
4120 AACCGCTACCGCGCGCGCATTAAGCAGATTAATCAATCAACCGCGCA 4169
||| .....: |||
815 ValLeuGlyArgLeuGlyLeuGlyValGlyLysArgIleGluLeuAlaG 831
4170 ACACATTTCATCAGCGCTTATTGAGCCCTGCTTACCGATCGCGCTT 4219
: .....: |||
831 yGlyArgGlnValGlnProTyrIleLysAlaSerValLeuGlnLupheA 848
4220 CCGGCAAAAGTCCGACGCGCGCATACCGCCTATTGGCGCAGATTC 4269
: .....: |||
848 spGlyArgGlyThrValArgThrAsnGlyIleAlaHisArgThrGluLeu 864
4270 GCGAAACCGCGCAGTGGCGGATGCGCGTAAACGCGAAATCAAGCTT 4319
||||| .....: |||
865 ArgGlyThrArg...AlaGluLeuGly..... 872
4320 CAGCGTGCCTCCAGCGCTGCGCGCGCAAGCGCGCAATTGGAAGCG. 4368
||||| .....: |||
873 .....LeuGlyMetAlaAlaAlaLeuGlyArgGlyHisSerLeuTyrAla 888
4369 .....CAGCAGACGCGCGCATCAATTAAGCTACCGCTGG 4404
: .....: |||
888 eTyrGlyTyrSerIleGlyProLysLeuAlaMetProIrrp 901
seq_name: SwissProt_40:PERT_BORPE
seq_documentation_block:
ID PERT_BORPE STANDARD; PRT; 910 AA.
AC P14283;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DE 01-OCT-1996 (Rel. 34, Last annotation update)
DE Pertactin precursor (Outer membrane protein P. 69). (P. 93).
GN PRN OR OMP69A.
OS Bordetella pertussis.
OC Bacteria; Proteobacteria; beta subdivision; Alcaligenaceae;
OC Bordetella.
OX NCBI_TaxID=520;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN-ISOLATE CN2992;
RX MEDLINE=89264462; PubMed=2542937;
RA Charles I.G., Dougan G., Pickard D., Chatfield S., Smith M.,
RA Novotny P., Morrissey P., Fairweather N.F.,
RT "Molecular cloning and characterization of protective outer membrane
RT protein P.69 from Bordetella pertussis.";
RL Proc. Natl. Acad. Sci. U.S.A. 86:3554-3558(1989).
RN [2]
RP REVISIONS TO 264 AND 332.
RX MEDLINE=92407514; PubMed=1527510;

```

```

RA Li J.L., Fairweather N.F., Novotny P., Dougan G., Charles I.G.;
RT "Cloning, nucleotide sequence and heterologous expression of the
RT protective outer-membrane protein P.68 pertactin from Bordetella
RT bronchiseptica.";
RL J. Gen. Microbiol. 138:1697-1705(1992).
RN [3]
RP X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS).
RX MEDLINE=96196517; PubMed=8609998;
RA Emsley P., Charles I.G., Fairweather N.F., Isaacs N.W.;
RT "Structure of Bordetella pertussis virulence factor P.69 pertactin.";
RL Nature 381:90-92(1996).
CC -! FUNCTION: AGGLUTININ THAT BINDS TO EUKARYOTIC CELLS; A PROCESS
CC MEDIATED BY THE R-G-D SEQUENCE. PERTACTIN MAY HAVE A ROLE IN
CC BACTERIAL ADHESION, AND THUS PLAY A ROLE IN VIRULENCE.
CC -! SUBUNIT: MONOMER.
CC -! SUBCELLULAR LOCATION: Outer membrane.
CC -! DISEASE: MAY CONTRIBUTE TO THE DISEASE STATE OF WHOOPING COUGH.
CC -! MISCELLANEOUS: SYNTHESIZED ONLY IN THE PRESENCE OF LOW Mg(2+)
CC CONCENTRATION.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: J04560; AAA22980.1; ALT_SEQ.
DR PIR: A32560; A32560.
DR InterPro: IPR003992; pertactin.
DR InterPro: IPR003991; pertactin-related.
DR PRINTS: PR01482; PERTACTIN.
DR PRINTS: PR01484; PERTACTINFAMLY.
KW Outer membrane; Signal; Virulence; Repeat.
FT SIGNAL 1 34
FT CHAIN 35 910
FT CHAIN 35 711
FT PROPE 712 910
FT SITE 260 262
FT FT
FT FT
FT DOMAIN 266 290
FT REPEAT 266 270
FT REPEAT 271 275
FT REPEAT 276 280
FT REPEAT 281 285
FT REPEAT 286 290
FT DOMAIN 579 593
SQ SEQUENCE 910 AA; 93452 MW; A169871E20A2E7DB CRC64;

alignment_scores:
Quality: 270.00 Length: 1156
Ratio: 0.554 Gaps: 53
Percent Similarity: 42.128 Percent Identity: 21.367

alignment_block:
US-09-303-518D-653 x PERT_BORPE ..
Align seg 1/1 to: PERT_BORPE from: 1 to: 910

1366 AACAGCCGCTGTCGCAAAATCGCAAGC.....ACGCT 1400
||||| .....||||| |||
2 AsnMetSerLeuSerArgIleValLysAlaAlaProLeuArgArgThr 18
1401 GCTGTCAGGCGCAAGG..... 1419
||||| .....|||||
18 rLeuAlaMetAlaLeuGlyAlaLeuGlyAlaAlaProAlaAlaHisAlaA 35
1420 .....GAAACCAAGGCTCGTCAGCGCGCAGGTAAGTCATCTTA 1464
||||| .....||||| |||
35 spTrpAsnAsnGlnSerIleValLysThrGlyLupArgGlnHisGlyIle 51

```

```
1465 GATCAGCAGCGCGAGATCAAGCAAAAAAACACCTTTAGTCAATCGG 1514
      |||:::|||||
      HisIleGlnSerAspProGlyValArgThrAlaSerGlyThr 68
1515 CTGG...GTACGGGAGGGGAGCGTGCACATGCAATGCCGATATGAGT 1561
      :|||:::|||||
      68 rIleGlyValSerGlyArgGlnAlaGlnGlyIleLeuLeuGlnAsn... 83
1562 TCAACCCCGCAACCTATTTCCGCTTTCGCGCGGAGCGTGGATTTG 1611
      :|||:::|||||
      84 .....ProAlaAlaGlnLeuGlnPheArgAsnGlySerValThrSer 97
1612 AACGGGCAATTCGCTTTCCTCCACCGCATTCAAATACCGATGAAGGCG 1661
      :|||:::|||||
      98 SerGly.....GlnLeuSerAspAspGlyIle 106
1662 GATGATTTGTCAACCAATCAAGCAAAAAAGATCCACGTTACCTACAG 1711
      :|||:::|||||
      106 e.....ArgArgPheLeuGlyThrValThrValLysa 117
1712 GCATTTAA.....GATATTACTACACCGGCAATTAACACACTTG 1752
      :|||:::|||||
      117 IagIlyLysLeuValAlaAspHisAlaThrLeuAlaAsnValGlyAspThr 133
1753 GATGCAAAAAAGAAATTCCTACACGGTTGTTGGCGGAAAGATGC 1802
      :|||:::|||||
      134 .....TrrAspAspAspGlyIleAl 140
1803 AACCAAAACGACGCGCGCTCAATCTGMAATTACCAACCGGAGAGACGG 1852
      :|||:::|||||
      140 AleuThrValAlaGlyGln.....GlnAlaGlnAlaSerIleAla 154
1853 ATCGCACTTACTGCTTCGCGCGGACAAATTA.....AACGGCAAT 1896
      :|||:::|||||
      154 spSerThrLeuGlnGlyAlaGlyValGlnIleGluArgIlyAlaAsn 170
1897 ATCAGCAAAACAAACGCGCAACTGTTTTCAGCGGACAGACGACGCA 1946
      :|||:::|||||
      171 ValThrValGlnArgSerAlaIleValAspGlyLeu..... 183
1947 CGCCTACATCATTTAGGAGCGGGTGTCA...AAAATGGAAGTATCC 1993
      :|||:::|||||
      184 .....HisIleGlyAlaLeuGlnSerLeuGlnProGlnAspLeuP 197
1994 CACAAGAGAAATCGTGTGGACAAAGATGGATGACCGGACA..... 2037
      :|||:::|||||
      197 rOProSerArgValAlaLeuArgAspThrAsnValThrAlaValProAla 213
2038 .....TTTAAAGC 2045
214 SerGlyAlaProAlaIleValSerValLeuGlyAlaSerGlnLeuThr 230
2046 GGAAGAACTTCATATTCAAGGCGGACAAAGCGGTGTTCCCGCAATGTTG 2095
      :|||:::|||||
      230 uAspGlyGlnHisIleThrGlnGlyArgAlaAla.....GlyValAla 244
2096 CCAAGTGAAGCGGATTCGATTAAACAATCAGCCCAAGCAAGTTTTC 2145
      :|||:::|||||
      244 Ia.....AlaMetGln 247
2146 GGTGTGCAACCGCATCAAGCCACACAATCTGTACACGTTGAGAC..... 2190
      :|||:::|||||
      248 GlYAlaValAlaHisLeuGlnArgAlaThrIleArgArgGlyAspAlaPr 264
2190 ..... 2190
264 oAlaGlyAlaValAlaProGlyGlyAlaValAlaProGlyGlyAlaValAlaProG 281
2191 .....TGACGGGT 2199
281 IyGlyPheGlyProGlyGlyPheGlyProValLeuAspGlyTrrPlyrGly 297

2200 CTGACAAAGTTGTACCGAAAAAACCATTAACGCAATGAAGTATGCTTC 2249
      :|||:::|||||
      298 Val..... 298
2250 ATTGAGCAAGACCGACATCAGAGC...AATGTACGCTTGGCCATCAG 2296
      :|||:::|||||
      299 .....AspValSerGlySerSerValGlnLeuAlaGln... 309
2297 CTCATTTAAATGTCACAGGACTTGGCCACATGCAAGGCAATGTTGATCA 2346
      :|||:::|||||
      310 .....SerIleValGlnAlaProGlnLeuGlyAlaAlaIleArgVal 323
2347 GCGGAGACACCGCACTATACGTTACGGCAAC.....GCCACCAAA 2390
      :|||:::|||||
      324 GlYArgGlyAlaArgValThrValSerGlySerLeuSerAlaProH 340
2391 CGGCAACTCAGGCTGCTGGGCAATGCCAAGCAACATTTAATCAAGC... 2439
      :|||:::|||||
      340 sGlyAsnValIleGluThrGlyAlaArgArgPheAlaProGlnAla 357
2440 .....ACATTAACGGCAACATCGGCTTGGACAAATGCT 2475
      :|||:::|||||
      357 IAProLeuSerIleThrLeuGlnAlaGlyAlaHisAlaGlnIlyAla 373
2476 TCATTTAATCTAAGCAACACCGCGTACAAAGCGAGTGCAGCTTTC 2525
      :|||:::|||||
      374 Leu.....LeuThrArgValLeuProGluProValIlyLeuThrLeuTh 388
2526 CGACAAAGCTAAGCAAGTAAGCATTCGCCACATCAACGGCATATGCT 2575
      :|||:::|||||
      388 rGlyGlyAlaAspAla.....GlnGlyAspIleVal 398
2576 CCCTACCGCATTAAGCGATATTCATTTTGAACACCGGCTTACCGGA 2625
      :|||:::|||||
      398 AlaIaThr.....GlnLeuProSerIleProGly 407
2626 AAATACGCGCGCGCAAGATACGCAATTAACGAGCAAGGAAAG 2675
      :|||:::|||||
      408 ThrSerIleGlyProLeuAspValAlaLeu...AlaSerGlnAlaArgTr 423
2676 GACGCTGCGGTGGGCAAGATTAAGCAATTAACCTTGACACGCA 2725
      :|||:::|||||
      423 PThr...GlyAlaThrArgAlaValAspSerLeuSerIleAspAsnAlaTr 439
2726 CCATTACACTCAATTCGCTATTCAGACAGATGCGGACGCGCAACC 2775
      :|||:::|||||
      439 hTrpValMetThrAsp..... 444
2776 GGCAGTGGCGGACAGATGCGCGCGCGCGCTTCGCGCTTCCCTATATC 2825
      :|||:::|||||
      445 AsnSerAsnValGlyAlaLeuArgAlaSerAspGlySer...ValAs 460
2826 CGTTACGCCCGCAACTTGGCAGCAATCCGTTTCAACACGCTGACGGTAA 2875
      :|||:::|||||
      460 PheGlnGlnProAlaGlnAla...GlyArgPheLysValLeuThrValAla 476
2876 ACGGCAAAATGAACGGTCAAGGAAATTCGCTTATGTGCGAACTCTTC 2925
      :|||:::|||||
      476 sn...ThrLeuAlaGlySerGlyLeuPheArgMetAsnValPheAlaAsp 491
2926 GCGTACCGGCGGCAATTTGAAGCTGGCGGAAAGTTCCGAAGCAGCTTA 2975
      :|||:::|||||
      492 LeuGlyLeuSerAspLysLeuValValMetGlnAlaSerIleGlnH 508
2976 CACCTTGGTGTCAACATACCGGCAAGCAAGCCGTAAGTCTTCGACCAAT 3025
      :|||:::|||||
      508 sArgLeuTrpValArgAsnSerGlySerGluProAlaSerAlaAsnThrL 525
3026 TGACGGTAGTGGAAGAAAAGACAACACACCGCTGTCCGAAATCTTAAT 3075
      :|||:::|||||
      525 euLeuLeuValGln.....ThrProLeuGlySerAlaAlaThr 537
3076 TTCACCTGCAAAACGAA.....CACGTGATGCGGCGCAATGGGCTTA 3119
```

```

|||||  |||||  |||||  |||||  |||||
538 PheThrLeuAlaAsnLysAspGlyLysValAspIleGlyThrTyrArgTyr 554
3120 TCAGCTTATCCGCAAGACGGCGAGTTCCGCCCTGCATTAATCCGGTCAAG 3169
554 rArgLeu..... 556
3170 AACAGAGCTTTCCGCAAACTCGCAAGGGGAGGAAACAGAGGCCGCC 3219
556 ..... 556
3220 TTACCGGCAAAACAGCAACTTCCGCCCAACAGCGGAAAGA 3269
557 ..... 560
3270 CAAGCGCAAAAGCCTTGAGCGCGTATGCGCGCGCGCGCAATGCCACCG 3319
|||||  |||||  |||||  |||||  |||||
560 yAsnGlyIntPr.....SerLeuValGlyAla..... 569
3320 AAAAGCAGAAAGTGTGGCGAAGCGCGCGCGAGCGGCGGAAAT 3369
|||||  |||||  |||||  |||||  |||||
570 ..LysAlaProProAlaProLysProAlaProGlnProGln... 584
3370 GCGCGATTATGACGGCGAGAGAGAAAAAGGGTGCAGCGCGATTA 3419
584 ..... 584
3420 AGACACCGCCTTGCGAAACAGCGGAAACCGCGCGCGCTACCA 3469
585 .....ProGlnPro..... 588
3470 CCGCCTTCCCGCGCGCGCGCGCGCGGATTTGCGCAACCGAG 3519
589 .....ProGlnProGln 592
3520 CCCCA.....CCGCAACCCCAACCGCGCGCGCATATACCG 3560
|||||  |||||  |||||  |||||  |||||
593 ProGlnAlaProAlaProGlnProAlaGlyArgGlyLeu...SerAl 608
3561 TTATGCCAATAGCGGTTGAGTGAATTTTCCGCAAGCTCAACAGCTT 3609
|||||  |||||  |||||  |||||  |||||
608 aAlaAlaAsnAlaAlaValAsnThrGlyValGlyLeuAlaSerThr 625
3610 ....TTGCGCGTACAGAGCAATGGACCGCGGTTCGCGAAGACCG 3654
|||||  |||||  |||||  |||||  |||||
625 eutPrTyAlaGlnSerAsnAlaLeuSerLysArgLeuGlyGlyLeuArg 641
3655 CGCAAC.....GCCGTTGGACAAGCGCGCATCCGGACACCA 3692
|||||  |||||  |||||  |||||  |||||
642 LeuAsnProAspAlaGlyAlaThrPrlArgGlyPhe..... 654
3693 ACCTACCGCTTCGCAAGATTTCCGCGCTACCGCGCAACACGACTGC 3742
655 .....AlaGlnArgGlnGlnLeuAsp... 661
3743 GCCAATCGGTATGCAAAAAACCTCGGACGGCGCGGTGGCATCTCG 3792
661 ..... 661
3793 TTTTCGCAACCGCGCGCAACACTTCGACGAC..... 3828
|||||  |||||  |||||  |||||  |||||
662 .....AsnArgAlaGlyArgArgPheSerGlnLysValAlaGlyPhe 675
3829 .....GGCATCGCAACTCGGACGGCGGTTCGCGACGGTCC.....G 3865
|||||  |||||  |||||  |||||  |||||
675 eGluLeuGlyAlaAspHisAlaValAlaValAlaGlyGlyArgTrpHisL 692
3866 TTTTCGGCAATACGGCATCGGAGGTTCCGATCGCATCGCATACGGCGGC 3915
|||||  |||||  |||||  |||||  |||||
692 euGlyGlyLeuAlaGlyTyrThrArgGlyAspArgGlyPheThrGlyAsp 708
3916 GCGGCTTTTACTAGCGGACGCTT..... 3939
|||||  |||||  |||||  |||||  |||||

```

```

709 GlyGlyGlyHisThrAspSerValHisValGlyGlyTyrAlaThrTyrI 725
3939 ..... 3939
725 eAlaAspSerGlyPheTyrLeuAspAlaThrLeuArgLaserArgLeuG 742
3940 .....TCAGACGCG.....ATCAGAGCAAA 3960
742 LuAsnAspPheLysValAlaGlySerAspGlyTyrAlaValLysGlyLys 758
3961 ATCCGCCCGCGCTGCTCATTCAGCATTCAGCAAGATACCGCGAG 4010
759 TyrArgThr.....HisGlyValGlyLaserLeuGlyAlaGly 771
4011 T.....TTGCGCGATTTCGCGCATCAACCGCACATCGCG 4045
771 yArgArgPheThrHisAlaAspGlyTyrPheLeuGlnProGlnAlaGlyL 788
4046 CAAGCGCGTATTGCTCCAAAAACCGGATTCGCGATACGAAAC...GTC 4092
788 euAlaValPheArgAlaGlyGlyGlyAlaTyrArgAlaAlaAsnGlyLeu 804
4093 AATATCGCCACCCCGCGCTTCATTCACCGCTACCGCGCGCATTA 4142
805 ArgValArgAspGlyGlyGlySerSerValLeuGlyArgLeuGlyLeu 821
4143 GCGAGATTATTCAATCAACCGCGCAACACATTTCATACCGCTTAT 4192
821 uValGlyLysArgIleGlyLeuAlaGlyArgGlnValGlnProTyrI 838
4193 TGACCGTCTCTATACGATCCGCTTCGCGCAAGTCGCAACCGCGCTC 4242
838 IeLysAlaSerValLeuGlnLupheAspGlyAlaGlyThrValHisThr 854
4243 AATACCCCGTATTGCGCAGGATTTGCGCAAAACCGCGATCGGATG 4292
855 AsnGlyLeuAlaHisArgThrGlnLeuArgGlyThrArg...AlaGlyLe 870
4293 GGGCGTAAACCGCAATCAAGTTTCACGCTGCTCCGTCACGCTGCCG 4342
870 uGly.....LeuGlyMetAlaAlaAlaL 878
4343 CCGGCAAGGGCGCAATTCGAACG.....CAGCAGACCGCGCATC 4386
878 euGlyArgGlyHisSerLeuTyrAlaSerTyrGlyLuprYserLysGlyPro 894
4387 AATTAGCTTACCGCTCG 4404
895 LysLeuAlaMetProIpr 900
seq_name: SwissProt_40:OMP_B_RICTY
seq_documentation_block:
ID OMP_B_RICTY STANDARD: PRT; 1645 AA.
AC P96989;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Outer membrane protein B precursor (168 kDa surface-layer protein)
DE (Surface protein antigen) (Cell surface antigen 5) (Scas) (OMPb)
DE (OMP B) (Contains: 120 kDa surface-exposed protein (Surface protein
antigen) (120 kDa outer membrane protein ompB); 32 kDa beta peptide).
GN OMP OR SLP.
OS Rickettsia typhi.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
OX NCBI_TaxID=785;
RN [1]
RC STRAIN=WILMINGTON.
RX MEDLINE=94040787; PubMed=8224886;
RA Hahn M.-J., Kim K.-K., Kim I., Chang W.-H.;
RT "Cloning and sequence analysis of the gene encoding the crystalline

```

RT surface layer protein of Rickettsia typhi.";
 RL Gene 133:129-133(1993).
 RN [2]
 RP PARTIAL SEQUENCE.
 RC STRAIN=WILMINGTON;
 RX MEDLINE=92114896; PubMed=1370573;
 RA Ching W.M., Carl M., Dasch G.A.;
 RT Mapping of monoclonal antibody binding sites on CNBR fragments of
 the S-layer protein antigens of Rickettsia typhi and Rickettsia
 prowazekii.";
 RL Mol. Immunol. 29:95-105(1992).
 RN [3]
 RP IDENTIFICATION OF CLEAVAGE SITE.
 RX Hackstadt T., Messer R., Cieplik W., Peacock M.G.;
 MEDLINE=92104668; PubMed=1729180;
 RT "Evidence for proteolytic cleavage of the 120-kilodalton outer
 membrane protein of rickettsiae: identification of an avirulent
 mutant deficient in processing.";
 RL Infect. Immun. 60:159-165(1992).
 CC -1- FUNCTION: THE 120 KDA SURFACE-EXPOSED PROTEIN IS A MAJOR
 STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS A RICKETTSIAL
 VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION.
 CC -1- FUNCTION: THE 32 KDA BETA PEPTIDE MAY SERVE AS A MEMBRANE ANCHOR.
 CC -1- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A
 S-LAYER WITH HEXAGONAL SYMMETRY.
 CC -1- SIMILARITY: BELONGS TO THE RICKETTSIAE OMPA/OMPB FAMILY.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: L04661; AAB48987.1.
 DR InterPro: IPR003858; romPA_romPB; 1.
 DR Pfam: PF02708; romPA_romPB; 1.
 KM Antigen: S-layer; Transmembrane; Cell wall.
 FT CHAIN 1 1353 120 KDA SURFACE-EXPOSED PROTEIN.
 FT TRANSMEM 1354 1645 32 KDA BETA PEPTIDE.
 FT TRANSMEM 1415 1429 MEMBRANE ANCHOR (POTENTIAL).
 FT CONFLICT 657 657 H -> N (IN REF. 2).
 FT CONFLICT 842 842 V -> I (IN REF. 2).
 FT CONFLICT 1071 1071 G -> A (IN REF. 2).
 FT CONFLICT 1306 1306 G -> S (IN REF. 2).
 SQ SEQUENCE 1645 AA; 169698 MW; 0CB5641C7EB185DE CRC64;

alignment_scores:
 Quality: 267.50 Length: 1683
 Ratio: 0.350 Gaps: 88
 Percent Similarity: 45.455 Percent Identity: 19.311

alignment_block:
 US-09-303-518D-653 x OMPB_RICTY ..
 Align seg 1/1 to: OMPB_RICTY from: 1 to: 1645

```

190 TTTGAGTGGGGGGAAGATATTAGTTTACAAACAAGAGGAGATT 239
    |||:|||||
125 PHERITILEGLYALA..... 129
240 GGTCCGAATGATGACGAAGAGCCCGCATGATGATTTTCTGTGAT 289
    |||:|||||
130 ...Glyylserleuthr.....Ilethrglyhslglylet 141
290 CGCGTAAGGCGCTGGCGCATTTGGCGGCGATCAATATATGTGACGCTG 339
    ::::|:|:|:|:|
141 hralaglnalalathrThrlyserAlaglnasnvalvalserlys 157
340 GCACATAACGGC.....GGCTATAACAA 362
    ::::|:|

```

```

158 valasnalaglYAlaAlaIleasnaspnaspleSerGlyValGlySe 174
363 TGTGATTTTGGTGGCGAGGAGAC.....AATC 391
    ::|:|:|:|:|
174 rliaspheThrAlaAlaProSerValleuIurheasnleuIleasnP 191
392 CGCATGACGACCGCTTTCTTACCAAAATGTGAAAAAATAATTAATAA 441
    ||:|:|:|:|
191 rothrThrGlnGlnAlaProleuThrleuGlyAspAsnalalysIleVal 207
442 GCAGGACTAAGCGCCATCTTATGCGGCATTTATCATGCCGCGTT 491
    |||:|:|:|:|
208 AsnGlyAlaAsnGlyIleleu..... 214
492 GCACAATTTGTCACAGATGCAGAACCTGTGAGATGACCACTATATANG 541
    |||:|:|:|:|
215 ....AsnIleThrAsn.....GlyPheVallys 222
542 ANGGGTGAATAATCGCTGATTAATAATACCTGATCGTTGGAATC 591
    ::::|:|:|:|:|
223 ValSerAspIlyThrPheAlaGlyIleIysThrIleAsnIle..... 236
592 GAGACGACGACAAATATGCGGCTGTGTAAGAACCCAAATATACCG 641
    ::|:|:|:|:|
237 .....GlyAspAsnGlnGlyleuMetPheasnThrProaspAla 251
642 CGAAATTCATATATATGCAAGCGCATATTTGGCTGTGCTGTGCGCA 691
    ||:|:|:|:|
251 la..AsnalaleuasnleuGlnGly.....GlyGlyA 261
692 ATACCTTTGCACAAATGCA...TCAGGTGGTGGCACACTCACTTAGT 738
    |||:|:|:|:|
261 snThrIleasnPheasnGlyArgAspGlyThrGlyLeuValleuVal 277
739 AGCGAAAAAATTAACATAGCCCATATGTTTATACCAACGAGAGCTC 788
    |||:|:|:|:|
278 SerIlyAsnGlyAsnalathrGluPheasn.....ValThrGlySe 291
789 ATTTGGCGAGAGTGGCTCACCAATGTTTATTCATGAT.....GCCCAA 832
    ::|:|:|:|:|
291 rleuGlnGlyAsnleuIysGlyValIleGluPheThrAlaAla 308
833 AGCAAAAGTGTAAATTAATGGGTATTCGCAACAGCAACCCCTATATA 882
    ||:|:|:|:|
308 laGlyIysleuIleAlaAsnGly.....GlyAlaIleAsnAlaValIle 322
883 GGAAGAAAGCAATGGCTCCAGCTAGTTCGTAAGATTTGTTATGATGA 932
    |||:|:|:|:|
323 GlyThrAspAsnGly.....AlaGlyArgAlaAlaGlyPheIleValSe 337
933 AATCTTTGCTGAGATACCATTCAGTATTTACGACCAACATCAAAATG 982
    ::::|:|:|:|:|
337 rValAspAsnGlyAsnalalathrIle.....SerG 348
983 GGAATACCTTTTAAAGCAGC.....AATAATATGCGCGCAGGA 1020
    |||:|:|:|:|
348 lyeIlnvalIyAlaIlyAspIleValIleGlnSerAlaAsnalaglyGly 364
1021 AAAATGCATGCCAACAATAAACATATCTCTACCTTATAGATTAANAAC 1070
    ::::|:|:|:|
365 GluValThrPheGluHis..... 370
1071 ACGAACCGTTCATTTGTTATATGTTCTTATCCGAGACAGCAAGAGAAC 1120
    |||:|:|:|:|
371 .....LeuValAspValGlyleu..... 376
1121 CTGTATATCATGCTGCAGGTGGGCTCACACAGTATGCA.....CCC 1161
    |||:|:|:|:|
377 .....GlyGlyIlyThrAsnheIysThrAlaAspSer 387
1162 AGACTGAATATGAGAGAAAATATTTCTTATGCAAGAGAAAGCTGA 1211
    ::::|:|:|:|:|
388 lyeValIleIleThrGluAsnalaserPheGlySerThrAspPheGlyAs 404

```



```
946 uTyrAlaGlyIleArgThrThrAsn..... 956
2552 ATTCCGACATCAAGCGCATGTCCTCCCTAGCCGATAGCAGTATTCAT 2601
957 .....GlnGlyThrValThrLeu..... 962
2602 TTGAAAAACAGCCGCTTACCGGAAAAATCAAGCGCGCAGATACG.. 2649
963 .....SerGlyGlyIleProAsnAspProGlyThrIle 973
2650 ....GCATTACCTTAAGACAGCGGAATGAGCGTCGCGCGGACG.. 2694
973 eTyrGlyLeuGlyLeuGlnAspProGlyLeuGlnValThr 990
2695 .....GAATTAGCAATTAAACCTTGACAAAGCCGACC 2727
990 heThrThrAspTyrAsnAsnLeuGlySerIleIleAlaThrAsnValThr 1006
2728 ATTACACTCAATTCGCCCTATCGACAGATGCGGCGCGCAACCGG 2777
1007 Ile.....AsnAspAspValThrLeuThrIleGln 1016
2778 CAGTGGCGCAGATGCGCGCGCGCGCTGCGCGCTTCCCTATTATCG 2827
1016 yGlyIleAlaGly..... 1020
2828 TTACGCGCCCAACTTCGGCAGATCCGTTTCAACAGCGTGACGTTAAC 2877
1021 .....ThrAspPheAspGlyLysIleThrLeu 1029
2878 GGCAATTGACGCGTCAAGGAAACATTCGCTTATGCGAACCTTCGG 2927
1030 GlySerIleAsnGlnAlaAsnValLysPheValAspArgThrPhe 1046
2928 TTACCGCGCGCAAAATTGAAGCTGCGGAAAGTTCCGAAGCAGCTTACA 2977
1046 rHisProThrSerMetIleValSerThrLysAlaAsnGlnIleThrVal 1063
2978 CC.....TTGGCTGTCAACATACCGCGACAGAA.....CCG 3009
1063 hrTyrLeuGlyAsnAlaLeuValGlyAsnIleGlySerSerAspIleThr 1079
3010 GTAACTCTCGAGCAATGACGCTAGTAGAGAAAGAAACACACACCGCT 3059
1080 ValAlaSerValArgPheThr.....GlyAsnAspSer..... 1090
3060 GTCCGAAAATCTTAATTTCACCTCGCAAGACAGACAGCTGACGCGCG 3109
1091 .GlyValGlyLeuGlnGlyAsnIleHisSerGlnAsnIleAspPheGly 1107
3110 CATGGCGTTATCAGCTATCCGCAAGAC..... 3138
1107 hrTyrAsnLeuThrIleLeuAsnSerAspValIleLeuGlyGlyThr 1123
3139 .....GGCAGTTCGCGCTG..... 3153
1124 ThrAlaIleAsnGlyIleIleAspLeuThrAsnAsnLeuIlePheAl 1140
3154 .....CATATCCGCTCAAAAGACAGAGCTT 3181
1140 aaSnGlyThrSerThrTyrGlyAsnAsnThrSerLeuSerThrThrLeuAl 1157
3182 CCGCAAACTCGCAAGCGGAGAAACAGAGCGCGCTTGACGGCAAAA 3231
1157 snValSerAsnGlyAsnValGlyGln.....IleValIleAlaGln 1170
3232 CAGGCAACAATTGCGCGCAACACAGCGCGAA.....AA 3266
1171 GlyAlaGlnValAsnAlaThrThrThrGlyThrThrIleLysIleGln 1187
3267 AGACAAACGGCAAGC.....CTTGACGGCGTGAATGG 3298
|||||
```

```
1187 nAspAsnAlaAsnAlaAsnPheSerGlyThrGlnThrThrThrLeuIleG 1204
3299 CGCGCGCGCGCC.....AATGCCACCGAAAGCGGAAGTGTGCCGAA 3342
1204 lngIleGlyAlaArgPheAsnGlyThr.....LeuGlyAla 1215
3343 CGCGCGCGCGCAGCGCGGGAATAATCCGCAATTATGACGCGGAGGA 3392
1216 ProAsnPheAspValThrGlyAsnAsnIlePheVal.....LysTyr 1229
3393 AGAGAAAAAACGGTGCAGCGCGATTAAGACACCGCTTGCGGAACAGC 3442
1229 rGlnLeuIleArg...AspAlaAsnGlnAspTyrValLeuThrArgThrAl 1245
3443 CGGAGCGGAAACCGCGCGCTACACCGCTTCGCCGCGCGCGCGCGC 3492
1245 snAspAla...LeuAsnValValThrThrAlaValGlyAsnSerAlaIle 1260
3493 CGCGCGCGGAAATTGCGCGCAACCGCGCAACCGCAACCGCAACCGCA 3542
1261 AlaAsn.....AlaProGlyValIle 1267
3543 GCGCGACCTGATCAGCCCTTATGCCAATAGCGTTGATGTAATTTCGG 3592
1267 scIAsnIle.....AlaIleCysLeuGlnSerThrAspThrAla 1280
3593 CCACGCTCAACAGCGTTTCGCCGTACAGACGAAATTGAGC..... 3633
1280 lAlaIleTyrAsnAsnMetLeuLeuAlaLysAspSerSerAspValAlaThr 1296
3634 .....CGCGTGTTCGCGAAGACCGCGCAACCGCGTTTGACAAAGCG 3677
1297 PheIleGlyAlaIleAlaThrAspThrGlyAlaAlaValAlaThrValAs 1313
3678 CATCCGGGACACCAACACTACCGCTCGCAGATTCGCCGCTACCGCC 3727
1313 nLeuAsnAspThrGln.....LysThrGlnAspLeuGlyAsnArgL 1328
3728 AACAAACCGACTCGCCCAATCGTATGCAGAA.....AACCTGCGC 3771
1328 euGlyAla...LeuArgTyrLeuSerAsnSerGlyThrAlaAspValGly 1343
3772 AGCGGCGCGCTCGCATCCGTTTTCGCACAAACCGCGGAACACCTT 3821
1344 GlySerGlyThrGlyAlaValSerSerGlyAspGlnAlaIleAspGlnVa 1360
3822 CGACGACGCGATC..... 3834
1360 lSerTyrGlyValThrAlaLysProPheTyrAsnIleAlaGlnGlnAspL 1377
3835 .....GGCACTCGGCACGCTTGCACGCTGCCGCTT 3867
1377 yLysGlyGlyLeuAlaGlyTyrLysAlaLysThrAla...GlyValVal 1392
3868 TTGGGCGCAATACGGCAATCGCGCAGTTCGACATGCGATCAGCGCGCGC 3917
1393 ValGlyLeuAspThrLeuAlaAsnAspAsnLeuMetIleGlyAlaAlaIle 1409
3918 GGGT.....TTTAGTAGCGGAGCC 3937
1409 euGlyIleThrLysThrAspIleLysHisGlnAspTyrLysGlyAspL 1426
3938 TTTTCAGAC.....GGCATCAGAGCGCAAAATCCGCGCGCGCGTGCAT 3981
1426 yStrThrIleLysGlyLeuSerPheSerLeu..... 1436
3982 TAGCGATTCAGCGCAAGATACCGCGCAGTTCGCGC...CGATTGCGCAT 4028
1437 TyrGlyAlaGlnGlnLeuValLysAsnPhePheAlaGlnIleSerAlaIle 1453
4029 C.....GAACGCAATCGCGCAACGCGCTATTGCTC.... 4062
1453 ePheThrLeuAsnLysValLysSerLysSerGlnArgTyrPhePheAspAl 1470
|||||
```

4063CAAAAGCGATACCGATACGAAAGTC 4092
 1470 laaanglylmetlaasnlyslntlelaalaaglyasnlyrpsasnhle 1486
 4093 AATATCGCCACCCGGCGCTTGATCAACCGCTACCGCGGGCATTA 4142
 1487 Thr.....PheglylsanleuMe 1493
 4143 GCGAGATTATTCATCAACCGCGCAACACATTCATCAGCGCTATT 4192
 1493 tPheglylTyrAspTyrAsnAlaLeuGlnGlyValLeuValThrPheMe 1510
 4193 TGAGCGTGTCTATACCGATCGCGCTCCGCAAGTCGCAAGCGCGTC 4242
 1510 laaglyleuSerTyrLeuLysSerSerAsnLysnlyrLysGluThgly 1526
 4243 AATACCGCCGATTTGGCGAGATTTCGCAAA...ACCGCAGTGGCA 4289
 1527 ThrThrValAlaAsnLysArgIleHisSerLysPheSerAspArgIleAs 1543
 4290 ATGGCGCGTAACGCGCAATCAAGTTTCAGCGTGTCCCTCCAC 4335
 1543 pleuilevalGlyAlaLysValThrGlySerAlaMetAsnIleAsn 1558

seq_name: SwissProt_40: OMPA_RICRI

seq_documentation_block:
 ID OMPA_RICRI STANDARD; PRT; 2249 AA.
 AC P15921:
 DT 01-APR-1990 (Rel. 14, Created)
 DT 01-APR-1990 (Rel. 14, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Outer membrane protein A precursor (190 kDa antigen) (Cell surface antigen) (r ompA) (r omp A).
 GN OMPA.
 OS Rickettsia rickettsii.
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
 OC Rickettsiaceae; Rickettsiae; Rickettsia.
 OX NCBI_TaxID=783;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-R:
 RX MEDLINE=90354033; PubMed=2117568;
 RA Anderson B.E., McDonald G.A., Jones D.C., Regnery R.L.;
 RT "A protective protein antigen of Rickettsia rickettsii has tandemly repeated, near-identical sequences";
 RL Infect. Immun. 58:2760-2769(1990).
 CC -1- FUNCTION: ELICITS PROTECTIVE IMMUNITY.
 CC -1- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A S-LAYER WITH HEXAGONAL SYMMETRY.
 CC -1- SIMILARITY: BELONGS TO THE RICKETTSIAE OMPA/OMPb FAMILY.
 CC -1- SIMILARITY: BELONGS TO THE RICKETTSIAE OMPA/OMPb FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC
 DR EMBL: M31227; AAA26380.1; -
 DR PIR: A41477; A41477.
 DR InterPro: IPR003858; r ompA_r ompb.
 DR Pfam: PF02708; r ompA_r ompb; 1.
 KW Antigen; Repeat; Signal; Cell wall; S-layer; Glycoprotein.
 FT SIGNAL 1 28 POTENTIAL.
 FT CHAIN 29 2249 OUTER MEMBRANE PROTEIN A.
 FT DOMAIN 212 1180 13 X APPROXIMATE TANDEM REPEATS.
 FT REPEAT 212 286 A (TYPE I).
 FT REPEAT 287 358 B (TYPE II).
 FT REPEAT 359 430 C (TYPE II).
 FT REPEAT

FT REPEAT 431 505 D (TYPE I).
 FT REPEAT 506 577 E (TYPE II).
 FT REPEAT 578 652 F (TYPE I).
 FT REPEAT 653 724 G (TYPE II).
 FT REPEAT 725 799 H (TYPE I).
 FT REPEAT 800 874 I (TYPE I).
 FT REPEAT 875 949 J (TYPE I).
 FT REPEAT 950 1021 K (TYPE II).
 FT REPEAT 1022 1093 L (TYPE II).
 FT REPEAT 1094 1165 M (TYPE II).
 FT REPEAT 1166 1180 TYPE I (INCOMPLETE).
 SQ SEQUENCE 2249 AA; 224333 MW; A9D646C089DF087 CRC64;

alignment_scores:
 Quality: 261.50 Length: 1472
 Ratio: 0.382 Gaps: 77
 Percent Similarity: 46.535 Percent Identity: 21.264

alignment_block:

US-09-303-518d-653 x OMPA_RICRI ..

Align seg 1/1 to: OMPA_RICRI from: 1 to: 2249

655 CATATTGCAAGCGCATATTCTTGCTGCTGGT.....GGCATACC.. 696
 938 AsnLeuAsnGlyAlaLeuSerGlnValThrGlyAspIleGlyAsnThrAs 954
 697TTTGCACAAAATGATCAGTGTGTCGACAGTCACTTAGTAGCG 742
 954 nserleuAlaThrIleSerValGlyAlaGlyThrAlaThrleuGlyGly 971
 743 AAAAAATTAAACATAGCCCATATGTTTACCAACAGAGCTCAATT 792
 971 laValIleLys.....AlaThrThrThrLysLeu 980
 793 GCGCAGAGGCTCA.....CCAATGTTATCTATGA 824
 991 ThrAspAlaIleSerAlaValLysPheThrAsnProValValThrGly 997
 825 TGCCCAAAAGCAAAAGTGTAAATTAATGG.....GTATGCAAAACAG 868
 997 yAlaIleAspAsnThrGlyAsnAlaAsnAsnGlyIleValThrPheThrG 1014
 869 GCAACCC.....TATATAGAAAACAAATGCTTCACACTA 906
 1014 LysAsnSerThrValThrGlyAsnValGlyAsnThrAsnAlaLeuAlaThr 1030
 907 GTTCGTAAAGATGTTCTATGATGAATCTTGCTGGA..... 945
 1031 ValAsnValGlyAlaGlyLeuLeuGlnValGlnGlyValValLysAl 1047
 946GATACCATCA.....GTATCTACGAAC 970
 1047 aAsnThrIleAsnLeuThrAspAsnAlaSerAlaValThrPheThrAsn 1064
 971 CA.....CATCAAAATGG 984
 1064 roValValValThrGlyAlaIleAspAsnThrGlyAsnAlaAsnGly 1080
 985 AATATCTTTTAAAGCAAT..... 1005
 1081 lleValThrPheThrGlyAsnSerThrValThrGlyAsnValGlyAsn 1097
 1006AATATGCGCAGCAAAATCGATCCAAAC 1036
 1097 rAsnAlaLeuAlaThrValAsnValGlyAlaGlyLeuGlnValGlnG 1114
 1037 ATAACACTATCTCTACCTTAGATTAATAACAGACCGTTCATATG 1086
 1114 LysGly.....ValValLysAlaAsnThrIleAsnLeu 1124
 1087 TTTAATGTTTCTTATTCGAGACAGCAAGAACTGTTATCATGCTGC 1136

```

1125 ThrAspSnaIaSerAlaValThrPheThrAsnProValVal...ValThr 1140
1137 AGGTGGGGTACACAGTTATCGACCCAGACATGATATGGAGAAATATTT 1186
1140 rGlyAlaIleAspAsn...ThrGlySnaIaAsnAsnGly...IleValThr 1155
1187 CCTTATT...GACAAAGGAAAGGTGAATTG 1215
1155 hrPheThrGlyAsnSerThrValThrGlyAsnIleGlyAsnThrAsnAla 1171
1216 ATACTTACGACCAATCAACCAAGC...GCGGCGGTTT 1253
1172 LeuAlaThrValAsnValGlyAlaGlyIleThrLeuGlnAlaGlyIse 1188
1254 GTATTGGAGGTAAATTTACGTCGCTTAAACACGAAAGGACGCGC 1303
1188 rLeuAlaIaAsnAsnIleAsnPheGlyAlaArgSer...ThrLeuG 1203
1304 AAGCGCGGCGGTCATATCATGATGACGATGACCTTACTTGGAAGTA 1353
1203 IupheAsnGlyProLeuAspGlyGlyGlySalaIleProTyrTyrPhe 1219
1354 AACGGC...GTGGCAACGACCGCTGTCCAAATCGGCAAGGACGCT 1400
1220 LysGlyAlaIleAlaAsnGlyAsnAsnAlaIleLeuAsnValAsnThrLy 1236
1401 GCTGTTCACGCAAGGGGAAACCAAGGCTCGTC... 1437
1236 sLeuLeuThrAlaSerHisLeuThrIleGlyThrValAlaGluIleAsnI 1253
1438 .....ACGTTGGCGGACGGTAAA 1455
1253 LeGlyAlaGlyAsnLeuPheThrIleAspAlaSerValGlyAspValThr 1269
1456 GTCACTTAGTACGACGCGGACGATCAAGCAAAACAAAGCCTTAG 1505
1270 ILeuAsnAlaGlnAsnIleAsnPheArgValaArgAspSerValIleVala 1286
1286 ILeuSerAsnLeuThrGlyValGlyValaAsnIleLeuLeuAlaIla 1303
1553 ATATCAATTC...AACCCGCAACAACCTATTTCCGCTTCGGGGCGGA 1599
1303 sPLeuValAlaProGlyAlaAspGlyThrValaIlePheAsnGlyGly 1319
1600 CGTTGGATTGAAAGCGGCAATTCGCTTCGATCCACGCAATTAAGATATAC 1649
1320 ValaAsnGlyLeuAsnValGlySerAsnValaIleGlyThrAlaArgAsnI 1336
1650 CGATGAAGGGCG...ATGATTGTCAACCA 1678
1336 eGlyAspGlyGlyAsnLysPheAsnThrLeuLeuIleTyrAsn... 1351
1679 ATCAAGACAAACATCCACCGTACCATTAACGCAATTAAGATATACT 1728
1352 .....AlaValThrIleThrAspAspValaAsnLeuIu 1362
1729 ACAACCGCAAT...ACAACAACCTGTATGCAAAAGAAAT 1769
1363 GlyIleGlnAsnValLeuIleAsnLysAsnAlaAspPheThrSerSerTh 1379
1770 TGCCCTACAAC...GGTTGGTTGGCGAGAAAGATCAACCAAAACG... 1812
1379 rAlaPheAsnAlaGlyAlaIleGlnIleAsnAspAlaThrTyrThrIleA 1396
1813 .....AACGGCGGCTCAATCTG... 1830
1396 sPAlaAsnAsnGlyAsnLeuAsnIleProAlaGlyAsnIleGlnPheAla 1412
1830 ..... 1830

1413 HisAlaAspAlaGlnLeuValleuGlnAsnSerSerGlyAsnAspArgTh 1429
1831 .....AATTACCAACCGGAAGAACGAT... 1854
1429 rIleThrLeuGlyAlaAsnIleAspProAspAsnAspArgGlyIleVal 1446
1855 .....CGCACTTACTCTGCTTCGGCGGA 1878
1446 alIleLeuAsnSerValThrAlaGlyLysLysLeuThrIleAlaGly 1462
1879 ACAAAATTAACGGCATATACGCAACCAACG... 1912
1463 LysThrPheGlyGlyAlaHisLysLeuGlnThrIleLeuPheLysGlyA 1479
1913 ..GCAAACTGTTTTCAGCGGACGACGACCGACCGCATCAATCAATT 1960
1479 lAglyAspCysSerThrAlaGlyThrThrPheAsnThrThrAsnIleVal 1495
1961 TAGGAAGCGGTGTCGCAAAATGAA...GCTATCCCAAGAGAGAAATC 2007
1496 LeuAspIleThrGlyGln...LeuGluLeuGlyAlaThrThrAlaAsnVal 1511
2008 GTGTGGACACAGATTCGATCGACCGCACATTTAAAGCGGA... 2049
1512 ValLeuPheAsnAspAlaValGlnLeuThrGlnThrGlyAsnIleGly 1528
2050 .....AACTTCATATTCAGCGCGGACACGCGGTGTTCCCGCATG 2092
1528 yPheLeuAspPheAsnAlaLysAsnGlyMetValThrLeuAsnAsnVal 1545
2093 TT...GCCAAGTGAAGCGGATTCGATTTAAGCAATCAGCC 2133
1545 alAsnValaIleGlyAlaValGlnAsnThrGlyThrAsnAsnGlyThr 1561
2134 CAAGCAGTTTCGTCGTCGACCGCATCAACCCACACATGTGACACG 2183
1562 LeuIleValleuGlyAlaSerAsn...LeuAsnAr 1572
2184 TTCGCACTGACGCGTCGTCGCAAGATTCGCAAAACCAATTAACGACG 2233
1572 yValaAsnGlyIleAlaMetLeuLysValaGlyAlaGlyAsnValThr... 1587
2234 ATAAAGTATTCGTTCAATGAGCAAGCCGACATC... 2268
1588 .....IleAlaLysGlyGlyValaLysIleGlyIuIleGlnGly 1601
2269 AGAGCAATGTACGCTTCGCGATCAGCGTCATTAATCTCAAGACAT 2318
1602 ThrGlyThrAsnThrLeuThrLeuProAlaHisPheAsnLeuThrGly 1618
2319 TGCACACCTCAACGCG...AATCTTAGTCAGGCG 2350
1618 rIleAsnLysThrGlyGlyGlnAlaLeuLysLeuAsnPheMetAsnGly 1635
2351 GAGACACGCACTATACGTTACGCGACGACGCCACCAAAACGCAACCTC 2400
1635 LysSerValSerGlyValaIleGlyThrAlaIaAsnSerValaGlyAspIle 1651
2401 AGCCTGCTGGC...AATGCCCAAGCAAC 2426
1652 ThrThrAlaGlyAlaThrSerPheAlaSerSerValaAsnAlaLysGlyTh 1668
2427 ATTTAATCAAGCCAAATTAACGCAACATCGGCTGCGAAGATGCTT 2476
1668 F...AlaThrLeuGlyGlyThrThrSer... 1676
2477 CATTTAATCTAAGCAACACGCGCTACAAAGCGAGTGTGACGCTTTC 2526
1677 .....PheAlaAsnThrPheThrAsnThrGlyAlaValThrLeu... 1689
2527 GACAACGCTAAGCAACGTAAGCATTCGCGACTCAACGCGCAATGTCTC 2576
1690 .....AlaLysGlySerIleThrSerPheAlaLysAsnValThrAlaTh 1704

```



```
2577 C.....CTAGCGATAGGCAAGTATTCATTTTGAAAAACAG...CGCT 2617
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1704 rSerPheValAlaAsnSerAlaThrIleAsnPheSerAsnSerLeuAlaP 1721
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2618 TTACCGGAAAAAATCAGCGCGGCAAGATAGCGCATTTACCTTAAAAAGC 2667
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1721 heAsnSerAsnIleThrIleGlyGly..... 1728
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2668 AGCGAATGAGCGCTGCGCGGCGGCAATTAGGCAATTTAAACCTTGA 2717
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1729 .....GlyThrThrLeu...ThrLeuGlyAlaAs 1737
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2718 CAACGCCACCATTTACATCATTCGCCCTATGACACAGATGGGAGGCG 2767
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1737 ngInValThrTyThr..... 1742
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2768 CGCAACCGCGCAGTGGCGGAGATGCGCGCGCGCGCTTGGCGCGTTC 2817
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1743 ..GlyThrGlySerPheThrAspThr.....Leu 1751
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2818 CTATATCCGTTAGCGCGCACTTCGGGAGATCCCGTTTCAACAGCT 2867
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1752 ThrLeuAsnThrThrPheAspGlyAlaIleAsnIleLe 1768
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2868 GACGGTAAACGGCAAA.....TTGAACGGTCAGGGAACATTCGCT 2908
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1768 uIleIysSerGlySerThrLeuAspLeuSerGlyValSerThrLeuAla 1785
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2909 TTATGTCGGAAGCTTTGGCTACCGCAGCGCAAAATGAGCTGGCGGA 2958
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1785 euValValThrAlaThrAsnPheAspMetAsnAsnIle..... 1797
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2959 AGTCCGAGGAGCACTTACACCTTGGCTGTCAACATACC.....GGCA 3002
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1798 SerProAspThrLysThrValIleSerAlaGluThrAlaGlyLe 1814
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3003 CGAACCCGTAAGTCTGAGCAATTCAGCGTAGTGAAGAAAAAGACACA 3052
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1814 uIysProThrSerLysGluAsnValLysIleThrIleAsnAsnAspAsn 1831
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3053 CACCGCTGCC.....GAAATCTTAAATTCACCCCTCGCAAAAGCA 3093
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1831 rgPheValAspPheThrPheAspAlaSerThrLeuThrPheAlaGlu 1847
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1848 AspIleAlaIleAspVal.....IleAspGlyAs 1857
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3094 CACGTCGATGCGGCGCATGGGTTATCAGCTTATCCGAAAGAGGCGGA 3143
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3144 GTTCGCGCTGCATTAATCCGTCAAAGAACAGCTTCCGACAAACTCG 3193
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1857 pPheAlaProGlyGlyProLeu..... 1864
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3194 GCAAGCGCGGAGAAACAGAGCGCGCTTGACGCAAAACAGGCAACTT 3243
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1865 .....AlaAsnIleProAsnAlaIleAsnIleLysIleSerLeuGlu 1879
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3244 GCGCGCAAAACACAGCGGAAAAAGACAACGCGCAAGCCTTGACGCGCT 3293
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1879 ..... 1879
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3294 GATTGCGCGCGGCGCAATGCCACGCAAAAGAGCAAAAGTGTGCGCAAC 3343
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1880 .....MetGluAspAlaProAsnGlySerAsp. 1888
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3344 GCGCGCGGAGGCGAGCGGGAATTCGCGGATATATGACGCGGAGGAA 3393
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1889 ..AlaArgGluAla...PheAsnAsnPheGlyLeuMetThrProLeuGlu 1903
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3394 GAGAAAAAAGCGGTCAGCGGATAAAGACACCGCTTGGCGAAACACGCG 3443
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1904 Glu.....AlaAspAlaThrThrIleLeu..... 1912
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3444 CGAAGCGGAAACCCGCGCGCTACACCGCTTCCCGCGCGCGCGCG 3493
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1912 ..... 1912
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3494 CCCCGCGGATTTGCCGCAACCGAGCCGCCAACCGCAACCCCAACGCGAG 3543
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1913 .....GlnAspValIleLysProSer..... 1919
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3544 GCGGACCTGATCAGCCGTTATGCCAATAGCGGTTTGAAGTATTTCCG 3593
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1920 ..AspThrIleAlaIleAlaIleAsnAsnGlnValAla.....SerAs 1933
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3594 CACGCTCAACAGCGCTTTCGCCGTACAGGAGCAATTAGACCGCGCTTTG 3643
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1933 nIleSerSerAsnIleThrAlaLeuAsnAlaArgMetAspLysValGln 1950
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3644 CC.....GAAGCCGCGCGC 3657
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1950 ergIAsnLysGlyProValSerSerGlyAspGluAspMetAspAlaLys 1966
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3658 AACCGCGTTTGACACAGCGC...ATCCGGAGACCAACAAACATACGCTTC 3704
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1967 PheGlyAlaThrPheSerProPheValGlyAsnAlaThrGlnLysMetC 1983
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3705 GCAAGATTTCCGCGCTACCGCCCAACCAACGAGCCTCGCCCAATCGCT 3753
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1983 sAsnSerIleSerGlyTyrLysSerAspThrThrGlyGlyThrIleGlyP 2000
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3754 .....ATGCAAGAAAAACCTCGGCAACGCGG..... 3777
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2000 heAspGlyPheValSerAspAspLeuAlaLeuGlyLeuAlaIleTyrThrArg 2016
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3778 ..CGCGTCGACATCTGTTTTCGACACACCGGAGCGGAAACACCTTGA 3824
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2017 AlaAspThrAspIleLysLeuLysAsnAsnLysThrIleLysPlys..... 2031
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3825 CGACGGCATCGCAACTCGGACAGCGCTTGGCCACGCGGCTTTCGGCG 3874
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2032 .....AsnLysValGluSerAsnIleTyrSerLeuTyrGlyL 2044
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3875 AATAGCGCATCGGAGGTTGACATCGGCAATCAACGCGGCGGCGGTTT 3924
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2044 euTyrAsnValProTyrGluAsnLeuPheValGluAlaIleAlaSerTyr 2060
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3925 AGTAGCGGACGCTTTCAGACGATCAGAGGCAAAATCCGCGCGCGCT 3974
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2061 SerAspAsnLys.....IleArgSerLysSerArgArgValIle 2073
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3975 GCTGCAATTAAGGCAATTCAGGCAAGATACCGCGAGGTTTCGGCGGATTCG 4024
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2073 eAlaThrThrLeuGluThrValGlyTyrGlnThrAlaAsnGlyLysTyrL 2090
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
4025 GCATCGAACCGCACATCGCGCAACGCGCTATTTCGTCAAAAAGCGGAT 4074
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2090 ysserGluSerTyrThrGlyGlnLeu.....MetAlaGly 2101
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
4075 TACCGATAC.....GAAACGTCAATATGCCACCGCGCGCGCTTGC 4115
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2102 TyrThrTyMetMetProGluAsnIleAsnLeu...ThrPro...LeuAl 2116
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
4116 ATTCAACCGCTACCGCGCGGCGCATTAAGGAGCATTAATTCATTAACCGG 4165
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2116 agLysLeuArgTyr...SerThrIleLys...AspLysGlyTyrLysGlu 2131
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
4166 CG.....CAACACATTTCCATTCACGCGCTTATTTGACCGCTGC 4203
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2131 hrGlyThrThrTyrGluAsnLeuThrValLysGlyLysAsnTyrAsnThr 2147
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
4204 TATACCGATCGCGCTTCGCGCAAGTCCGAACGCGGTCGAATACG..... 4248
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2148 PheAspLysLeuLeuGlyAlaLysValSerSerAsnIleAsnValGln 2164
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
4249 .....GCCGTATTGGCGCAGAGATTTCGCA 4273
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
```

```

|||||:
2164 uilleValLeuThrProGluLeuTyraLAlaMetValAspTyraLAlaPheLysA 2181
4274 AAMCCCGCATGCGGATGCGGCGTAACGCCGAATCAAGCTTCCACG 4323
      |||||
      .....IleAspAlaArgLeuGlnIlyMetThr 2194
2181 snLysValSerAla.....
4324 CAGTCCCTCCACGCTGCCGCGCAAGGCGCGCAATTGGAAGCGCACGA 4373
      |||
      .....GlnSerLysThrSerPh 2209
2195 AlaProLeuProThrAsnSerPheLys.....
4374 CAGCGCGGCGCATC 4386
      :
2209 eaSpValGlyVal 2213

seq_name: SwissProt_40:OMPA_RICCN

seq_documentation_block:
ID OMPA_RICCN STANDARD; PRT; 2021 AA.
AC Q52657; P95591; P95592; P95594; Q52667; Q52668; Q52669;
AD Q52670; Q52674;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Outer membrane protein A precursor (190 kDa antigen) (Cell surface
DE antigen) (OMP A) (OMP A).
GN OMPA OR R1273.
OS Rickettsia conorii.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
OX NCBI_TaxId=781;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Malish 7;
RX MEDLINE=94171067; PubMed=8125327;
RA Crocquet-Valdes P.A., Weiss K., Walker D.H.;
RT "Sequence analysis of the 190-kDa antigen-encoding gene of Rickettsia
RL conorii (Malish 7 strain).";
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Malish 7;
RX MEDLINE=21442074; PubMed=11557893;
RA Ogata H., Audic S., Renesto-Audiffren P., Fournier P.-E., Barde V.,
RA Sanson D., Roux V., Cossart P., Weissenbach J., Claverie J.-M.,
RA Raoult D.;
RT "Mechanisms of evolution in Rickettsia conorii and R. prowazekii.";
RL Science 293:2093-2098(2001).
RN [3]
RP SEQUENCE OF 8-204 FROM N.A.
RC STRAIN=Indian tick typhus, M1, Malish 7, and Moroccan;
RX MEDLINE=97015921; PubMed=8862558;
RA Roux V., Fournier P.E., Raoult D.;
RT "Differentiation of spotted fever group rickettsiae by sequencing and
RT analysis of restriction fragment length polymorphism of PCR-amplified
RT DNA of the gene encoding the protein ompA.";
RL J. Clin. Microbiol. 34:2058-2065(1996).
RN [4]
RP SEQUENCE OF 953-1012 FROM N.A.
RC STRAIN=Indian tick typhus, M1, Malish 7, and Moroccan;
RA Raoult D., Fournier P.E., Roux V.;
RT "Phylogenetic analysis of spotted fever group rickettsiae by study
RT of the outer surface protein ompA.";
RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.
CC -|- FUNCTION: ELICITS PROTECTIVE IMMUNITY (BY SIMILARITY).
CC -|- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A
CC -|- S-LAYER WITH HEXAGONAL SYMMETRY.
CC -|- PM: GLYCOSYLATED (BY SIMILARITY).
CC -|- SIMILARITY: BELONGS TO THE RICKETTSIAE OMPA/OMP B FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC at the European Bioinformatics Institute. There are no restrictions on its

```

CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).

```

DR EMBL; 001028; AAI17405.1; -
DR EMBL; AE008674; AAL03811.1; -
DR EMBL; U43794; AAB49549.1; -
DR EMBL; U43798; AAB49550.1; -
DR EMBL; U43806; AAB49551.1; -
DR EMBL; U45244; AAB49556.1; -
DR EMBL; U46918; AAB46663.1; -
DR EMBL; U83440; AAC35176.1; -
DR EMBL; U83443; AAC35179.1; -
DR EMBL; U83448; AAC35184.1; -
DR EMBL; U83453; AAC35189.1; -
DR InterPro: IPR003858; romPA_romPB.
DR Pfam: PF02708; romPA_romPB.1.
KW Antigen; Repeat; Signal; Cell wall; S-layer; Glycoprotein;
KW Complete proteome.
FT SIGNAL 1 38
FT CHAIN 39 2021
FT DOMAIN 238 946
FT DOMAIN 1424 1528
FT VARIANT 60 60
FT VARIANT 76 76
FT VARIANT 86 137
FT VARIANT 126 133
FT VARIANT 953 954
FT VARIANT 1245 1245
FT VARIANT 1308 1308
FT VARIANT 1877 1877
FT CONFLICT 10 10
FT CONFLICT 92 92
FT CONFLICT 126 126
FT CONFLICT 137 137
FT CONFLICT 157 157
FT CONFLICT 368 369
FT CONFLICT 374 388
FT CONFLICT 640 640
FT CONFLICT 669 669
FT CONFLICT 793 793
FT CONFLICT 804 804
FT CONFLICT 809 823
FT CONFLICT 898 898
FT CONFLICT 908 908
FT CONFLICT 985 985
FT CONFLICT 1009 1009
FT CONFLICT 1013 1013
FT CONFLICT 1182 1182
FT CONFLICT 1314 1314
FT CONFLICT 1451 1451
FT CONFLICT 1624 1624
FT CONFLICT 1628 1628
FT CONFLICT 1872 1872
FT CONFLICT 1875 1875
FT CONFLICT 1878 1879
FT CONFLICT 1936 1936
FT CONFLICT 1965 1970
FT CONFLICT 1997 1997
SO SEQUENCE 2021 AA; 203328 MW; 327FC42D7CB24668 CRC64;

```

alignment_scores: quality: 250.50 length: 1581
 ratio: 0.352 gaps: 78
 Percent Similarity: 45.035 Percent Identity: 20.304

alignment_block: US-09-303-518d-653 x OMPA_RICCN ..

Align seg 1/1 to: OMPA_RICCN from: 1 to: 2021

```
152 ACCAATACATATGCCGACTTGGCGAAATAAAGCAAGTTTGCAGTCGGG 201
      |||||||:|||||:
725 ThrAsnAlaLeuAlaThrIleSerValGlyAlaGlyLysAlaThrLeuG 741
202 GCGAAGATATGAGTTTACAAACAAAAGGAGTGGTCGGCAATC 250
      | :|||:|
741 yelAlaValIleLysAlaThrThrLysLeuThrAspAsnAlaSerA 758
251 G.....ATGACAAAGCCCGATGATTTCT 282
      :|||:|
758 lavalThrPheThrAsnProValValThrGlyAlaIleAspAsnThr 774
283 GTGGATCCCTAACGGCGTGGCGCATTTGGCGCATATATATGT 332
      :|||:|
775 GlyAsnAlaAsnAsnGlyIleAlaThrPheThrGlyAspSerThrValTh 791
333 G...AGCGTGCAATACGGCGCTATACAAATGTGATTTGGTGGG 379
      :|||:|
791 rGlyAsnIleGlyAsnThrAsnAlaLeuAlaThrValAsnValGlyAla 807
380 AGGGAAGCAATCCGATCAGACCGCTTTCTTACCAATTTGAAAGA 429
      :|||:|
808 .....GlyLeuLeuArgValGlnGlyValValLysSer 819
430 AATTAATTAAGCAAGGACTAACGGCCATCTTATGGCGCGATTAACA 479
      |||:|
820 AsnThrIleAsnLeuThrAspAsnAla..... 828
480 TATGCCGCTTTGCACAAATTTGTTCACAGATCAGAACCTTTGAGATGA 529
      :|||:|
829 .....SerAlaValThrPheThrAsnProValValValT 840
530 CCAGTTATATGATGATGGTGGAATACGCTATTAAATTAATCCCTGAT 579
      |||:|
840 hGlyAlaIleAsp..... 844
580 CGTGTTCGAATCGGAGCAAGCAGACATATTGGCGTGTGATGAAGAGA 629
      :|||:|
845 .....AsnThrGlyAs 848
630 ACCCAATAACCGCGAAGTTTCATATCATATTGCAAGCCATATTCTTGGC 679
      :|||:|
848 nAlaAsnAsnGlyIleValThrPheThrGlyAspSerThrValThrGlyA 865
680 TCGTGGGTGCG...AATACCTTGCACAAATGATCAGGTGGTGGCACA 726
      :|||:|
865 snIleGlyAsnThrAsnAlaLeuAlaThrIleSerValGlyAlaGlyLys 881
727 GTCAACTTGTAGCGAAATAAATTAACATACCCATATGTTTTTACC 776
      :|||:|
882 AlaThrLeuGlyGlyAlaIleLys.....Al 891
777 AACAGAGGCTCATTTGGCGACAGTGCCTCA.....C 808
      :|||:|
891 aThrThrThrLysLeuThrAspAsnAlaSerAlaValThrPheThrAsn 908
809 CATGTTTATCTATGATGCC...CAAAAGCAAAAGTGGTTAATTAATGCG 855
      :|||:|
908 rovalValValThrGlyAlaIleAspAsnThrGlyAsnAlaAsnAsnGly 924
856 GTATTGCAA.....ACAGCAACCCCTATATAGG 884
      :|||:|
925 lIleValThrPheThrGlyAspSerThrValThrGlyAsn.....IleG 939
885 AAAAAGCAATGGCTTCAGACTGATTCGTAAAGATGGTTCTATGATGAAA 934
      :|||:|
939 yAsnThrAsnAlaLeuAlaThrVal.....AsnV 949
935 TCTTTGCTGGAGATACCCATTCAGTATTTCTACGAACCATCAAAATGGG 984
      :|||:|
```

```
949 aLgLyAlaGlyValThr..... 954
985 AATATCTTTTAAACGACATATATATGCGCGAGAAAATCGATGCCAA 1034
      :|||:|
955 .....LeuGlnAlaGlySerLeuAspAlaAs 964
1035 ACATTAACACATATCTCTACTTATATATTAATAAACGACGACCTTCAT 1084
      :|||:|
964 nasIleAspPheGly.....AlaArgSerThrLeuG 975
1085 TGTTAATGTTCCTTATCCGAGACAGACAGAACCTGTTATCATGCT 1134
      :|||:|
975 lupheAsnGlyProLeuAsp..... 981
1135 GCAGGTGGGCTCAACACT.....TATCGACCCAGACTGAATAA 1172
      :|||:|
982 ..GlyGlyGlyAsnAlaIleProTyrTyrPheLysGlyAlaIleAlaAs 997
1173 TGGAGAAATATTTCTTATTTGACAAAGAAAGTGATTTGATCTTA 1222
      :|||:|
997 nGlyAsnAsnAlaIleLeuAsnValAsnThrLys.....LeuLeuT 1011
1223 CC.....AGCAATCAACCAAGCGCG 1245
      :|||:|
1011 hralATyrHisLeuThrIleGlyThrValAlaGluIleAsnIleGlyAla 1027
1246 GCGGCTTTGAT.....TTGAGGCTAATTTACGCTTCGCC 1283
      :|||:|
1028 GlyAsnLeuPheAlaIleAspAlaSerAlaGlyAspValThrIle..... 1042
1284 TAAAAACACGAACGTCGCAAGCGCGGCGCTCATATACGT.....G 1327
      :|||:|
1043 .....LeuAsnAlaGlnAspIleHisPheArgAlaLeuA 1054
1328 ATGCGACTACCTTACTTGGAAGTAACGCGCTGGCAACGACCGCTG 1377
      :|||:|
1054 sPserAlaLeuValLeuSerAsnLeuThrGlyAla..... 1065
1378 TCCAAATTCGCGCAAGCAGCGCTGCTGTCA.....GCCAA 1415
      :|||:|
1066 .....GlyValAsnAsnIleLeuLeuAlaAlaAspLeuValAlaPr 1079
1416 AGGGAACCAACGAGCTCGGTGCGGCGGCGAGCAAGTCAATCATCTAG 1465
      :|||:|
1079 oGlyValAspLeuGlyThrValAlaIleAspGlyGlyValAsnGlyLeuA 1096
1466 ATCAGCAGCGGCGAGATCAAGGCAAAACACGCTTACTGAATCGGC 1515
      :|||:|
1096 snIleGlySerAsnValAlaGlyAlaAlaArgAsnIleGlyAspValGly 1112
1516 TTGCTCAGCGGCGAGGAGCGGTGCACATGATCCGATATCATGTTCAA 1565
      :|||:|
1113 .....GlyAsnLysPheAs 1117
1566 CCCCAGCAAACTCATTTTCGGCTTCGCGGCGGACGTTGATTTGAACG 1615
      :|||:|
1117 nThrLeuLeuIleTyr..... 1122
1616 GGCATTGCGCTTTCGTCACCGCATTCAAATATCCGATGAAGGCGCATG 1665
1122 ..... 1122
1666 ATTGTCAAACCAATCAAGAACAAGATCCAGCTTACATTAACGGCAA 1715
      :|||:|
1123 .....AsnAlaValThrIleThrAspAs 1130
1716 TAAAGATATTACTACAACCGGCAAT.....AACCAACACTTGATA 1756
      :|||:|
1130 pValAsnLeuGlnGlyIleGlnAsnValLeuIleAsnAsnAlaAsp 1147
1757 GCAAAAAGAAATTTGCTACAAAC...GGTTGGTTTGGCGGAAAGATGCA 1803
      :|||:|
1147 heThrSerSerThrAlaPheAsnAlaGlyThrIleGlnIleAsnAspAla 1163
```

```

1804 ACCAAACG.....AACGGCGGCTCAATCTG.....AA 1832
1164 ThrTyrThrIleAspAlaAsnAsnGlnLeuLeuAsnIleProAlaGlyAs 1180
1833 TTACCAACCGGAAGAACGGATCGACTTTACTGCTTCGCGGGAACA 1882
1180 nIleYsPheAlaHisAlaSpAlaGlnLeuLeuGlnAsnSerSerg 1197
1883 ATTAAACGGCAATATCAGCAACAACGCAACTGTTTTCAGCGCG 1932
1197 LYAsnSPArGThrIleThr..... 1203
1933 AGACGACACCGCACCGCTTACATTCATTAGAGCGGCGTCAAAAT 1982
1204 .....LeuGlyAla..... 1206
1983 GGAAGGTATCCACAAAGAAATCGTGGCAACAGAT..... 2022
1207 .....AsnIleAspProAspAsnAspAspGlyGlyI 1217
2023 .....TGGATCGACCGCACATTTAAAGCGGAAACTTCATATTCAGGCG 2067
1217 leValIleLeuAsnSerValThrAlaGlyLysLysLeuThrIleAlaGly 1233
2068 GGACAAAGCGGTGGTTCCCGCAAT.....GTTGCCAAAGT 2102
1234 GlyLysThrPheGlyGlyAlaHisLysLeuGlnAspIleValPheYsG 1250
2103 GGAAGCGATTGGCATTTAAGCAATCAGCCCAAGCAGTTTTCGCTG 2152
1250 yGlyGlyAsp.....PheGlyThrA 1257
2153 CACCGCATCAAGCCACACATCTGTACAGCTTCGACGTGACGGGTG 2202
1257 laGlyThrThrPheAsnThrThrAsnIleValLeuAspIleThrGlyGln 1273
2203 .....ACAAGTTGTACGAAAAAACCATTTACCGAGATTAAGTATTC 2246
1274 LeuGluLeuGlyAlaThrThrAlaAsnValValLeuPheYsAspAlaVa 1290
2247 TTCATTGACCAAGACCGACATCAGAGGCAATGTC..AGCTTGGCGATC 2293
1290 LglnLeuThrGlnThr.....GlyAsnIleGlyGlyPheLeuAsp 1304
2294 AGCATATTAAATCTCACAGGACTGCCACATCACTCAAGCGCAATGTACT 2343
1304 heAsnAlaYsAsn.....GlyThrValThrLeuAsnAsnValAsn 1318
2344 GCAGGCGAGACACGACATATACGTTACGCGCAACGCCCAACCAACGG 2393
1319 ValAlaGlyThrValYsAsnThr.....GlyGlyThrAsnAsnG 1332
2394 CAACCTCACCGCTGGTGGCAATGCCCAACACATTTATCAAGCCACT 2443
1332 yThrLeuIleValLeuGlyAlaSer.....AsnLeuAsnArgV 1345
2444 TAAACGCAACACATCGCTCGACAAATCTTCATTTATCTTAAGCAAC 2493
1345 alaSnGly.....IleAlaMet 1350
2494 AACCGCTTCAAAACGGACGTGACGCTTCCGACACGCTAAGGCAAA 2543
1351 LeuYsValGlyAlaGlyAsnValThrIleAlaYsGlyYsValY 1367
2544 CGTAAACCGTCCGACCTCAAGCGCAATGCTTCAGCCGATTAAGCGAG 2593
1367 sTieGlyGluIleGlnGlyThrGlyThrAsnThrLeuThrLeuProAla 1383
2594 TATTCATTTGAAAAACAGCCGCTTACGGAATAATCAACGCGCGCAAG 2643
1384 .....HisPhe.....LysLeuThrGlySerIleAsnLysThrGly 1395
2644 GATACGGCATTTACCTTAAAGACAGCAATGACGCTGCCGCGGC.. 2691
1396 GlyAlaIleLeuLysLeuAsnPheLeuAsnGlyYsSerValSergYVa 1412
2692 .....ACGGAATTAGCAATTTAAACCTTGACACGCCCA 2725
1412 lValGlyThrAlaAlaAsnSerValGlyAspIleThrThrAlaGlyAlaT 1429
2726 CCATT..... 2730
1429 hrSerPheAlaSerSerValAsnAlaYsGlyThrAlaThrLeuGlyGly 1445
2731 ACATCAATTCGCCCTATCGACAGATCGCGACGCGCAAC... 2775
1446 ThrThrSerPheAlaHisThrPheThrAsnThrGlyAlaValThrLeuAl 1462
2776 ...GGCAGTGGCAGATGCGCGCGCGCGCTGCGCGCTTCCCTAT 2821
1462 aLYsGlySerIleThrSerPheAlaYsAsnValThrAlaThrSerPheV 1479
2822 TATCCGTTACCGCCCAACTCGCGCAGATCCGCTTC...AACAGCTG 2868
1479 AlaAla.....AsnSerAlaThrIleAsnPheGlyAsnSerLeu 1491
2869 AGGTTAAACGGCAATTTGAACGTCAAGGAACATTCGCTTATGTGGA 2918
1492 AlaPheAsnSerAsnIleThrGlySerGlyThrThrLeuThrLeu... 1506
2919 ACTCTGGGCTACCGCAGCGCAATTTGAAGCTGGCGGAAGTTCCGAG 2968
1507 .....GlyAlaAsnGlnValThrTyrThrGlyThrG 1517
2969 GCACCTAC.....ACCTTGACCTGCAACAATAC... 2997
1517 LysSerPheThrAspThrLeuThrLeuAsnThrThrPheAspGlyAlaAla 1533
2998 .....GGCAAGAACCCGTAAGTCTCGACCAATTTGACGTAAGTGA 3038
1534 LysSerGlyGlyAsnIleLeuIleLysSerGlySerThrLeuAspLeu 1550
3039 AGGAAGACACACACACCGCTGCCGAAT...CTTAATTTACCGCTGC 3085
1550 rGlyValSerAsnLeuAlaLeuValValThrAlaThrAsnPheAspPhe 1567
3086 AAAACGAACAGTGCATCGCGCATGCGCTTATCCGCAAA 3135
1567 snAsnIleSerProAspThr.....LysTyrThrValIle... 1578
3136 GACGCGAGTTCGCTGCATATCGGTCAAAGACAGAGCTTTCGA 3185
1578 ..... 1578
3186 CAACCTCGGCAAGCGGGAGAAACAGAGCGCGCTG...ACGGCAA 3229
1579 .....serAlaGluThrAlaGlyGlyLeuYsProThrProl 1591
3230 AACAGCAACAATTGCCGCCAAACACAGCGCGGAAAAAGAAC... 3273
1591 ySGLu.....AsnValYsIleThrIleAsnAsnAspAsnArgPhe 1604
3274 .....GGCAAGCCTTGACGCGCTGATTTGCGGCGG 3305
1605 ValAspPheThrPheAspAlaSerThrLeuThrLeuPhe..... 1617
3306 GCGCAATGCCACGAAAGAGCAAAAGTGTGCCGAACGCGCGGAGG 3355
1618 .....AlaGluSpIleAlaIleGlyValIleAspGlnPheAlaP 1632
3356 CAGCGGGGAAAAATCCGCGATTATGACGCGGAGGAAGAAAAACGG 3405
1632 roGlyGlyProLeuAlaAsnIleProAsnAlaAlaAsnIleYsLysSer 1648
3406 GTGCAAGCGGATTAAGACACC..... 3426

```

```

1649 LeuGluMetGluAspAlaProAsnGlySerAspAlaArgGlnAlaPhe 1665
      ::::: :::::
3427 .....GCCTTGGCGAAACAGCCGCGAAGCGGAAACCCGGCCGG 3463
      ::::: :::::
1665 eAsnAsnPheGlyLeuMetThrProLeuGlnGlnAlaAsp.....A 1679
3464 CTACACCGCGCTTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCA 3513
      |||||
1679 LeuThrThr.....HisLeuMetGlnAspValValLys 1689
3514 CCGGACCGCCCAACCGCAACCCCAACCGCAACCGCGCGCGCGCGCG 3563
      |||||
1690 ProSer.....AspThrIleAlaVala 1697
3564 TGCCAAATAGCGGTTTGAGTAAATTTCCGCGCGCGCGCGCGCGCG 3613
      |||||
1697 LAsnAsnGlnValaVala.....SerAsnIleSerSerAsnIleThrA 1712
3614 CCGTACAGAGCAATTTGACCGCGCGCGCGCGCGCGCGCGCGCGCG 3645
      |||||
1712 LAlaAsnAlaArgMetAspLysValaGlnAlaLysAlaLysGlyProVal 1728
3646 .....GAGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3677
1729 SerSerGlyAspGlnAspMetAspAlaLysPheGlyAlaThrPheSerPr 1745
3678 C...ATCCGGGACACCAACACTACCTGCGCGCGCGCGCGCGCGCG 3724
1745 opheValGlyAsnAlaThrGlnLysMetLysAsnSerIleSerGlyTyrL 1762
3725 GCCAACAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3774
      :::::
1762 ySerSerThrThrThrGlyGlyThrIleGlyPheAspLysPheValSerAsp 1778
3775 GGGCGCGGTC.....GGCATCTCTGT 3794
1779 AspLeuValLeuGlyLeuAlaTyrThrArgAlaAspThrAspIleLysLe 1795
3795 TTGGCAACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3844
1795 uLysAsnAsnLysThrGlyAspLys.....AsnLysV 1806
3845 CAGCGGTTCCCAACGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3894
1806 aGluSerAsnIleTyrSerLysThrGlyLeuTyrSerValProTyrGln 1822
3895 GACATCGGCGATCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3944
1823 AsnLeuPheValGlnAlaIleAlaSerTyrSerAspAsnLys..... 1836
3945 CCGCATCAGAGCGCAAAATCCCGCGCGCGCGCGCGCGCGCGCGCG 3994
1837 ...IleArgSerLysSerArgArgValIleAlaThrThrLeuGlnThr 1852
3995 CAAGTAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4044
1852 aGlyTyrGlnThrAlaAsnGlyLysTyrLysSerGluSerTyrThrGly 1868
4045 GCACGCGCGCTATTTCGTCACAAAGCGGATTAACGATAC.....GA 4085
1869 GlnLeu.....MetAlaGlyTyrThrTyrMetSerG 1880
4086 AAAGCTCAATATCGCACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4135
1880 uAsnIleAsnLeu...ThrPro...LeuAlaGlyLeuArgTyr...SerT 1894
4136 GCATTAAGCGAGTATTATTCACACCGCGCG.....CAACAC 4173
1894 hTleLys...AspLysSerTyrLysGlnThrGlyThrThrTyrGlnAsn 1909
4174 ATTTCATCAGCGCGCTATTATTCAGCGCGCGCGCGCGCGCGCGCG 4223
      :::::

```

```

1910 LeuThrValLysGlyLysAsnTyrAsnThrPheAspGlyLeuLeuGlyAl 1926
4224 CAAGTCCGAAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4248
1926 aLysValSerSerAsnIleAsnValAsnGlnIleValLeuThrProGlnL 1943
4249 .....GCCGTAATGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4293
1943 euTyrAlaMetValAspTyrAlaPheLysAsnLysValSerAla..... 1957
4294 GCGCTAAACCGCGCAATCAAGGTTTCACCGCGCGCGCGCGCGCGCG 4343
1958 ...IleAspAlaThrGlnGlnGlyMetThrAlaProLeuProThrAsnSe 1973
4344 CGCCAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4386
1973 rPheLys.....GlnSerLysThrSerPheAspValGlyVal 1985

```

seq_name: SwissProt_40:AIDA_ECOLI

seq_documentation_block:

```

ID AIDA_ECOLI STANDARD; PRT; 1286 AA.
AC 003155;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 01-JUN-1994 (Rel. 29, Last annotation update)
DE Adhesin aida-I precursor.
GN AIDA-I.
OS Escherichia coli.
OC Plasmid p186.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 50-56.
RC STRAIN=2787 (0126:H27);
RX MEDLINE=9232638; PubMed=1625582;
RA Benz I., Schmidt M.A.;
RT diarrhoeagenic Escherichia coli strain 2787 (0126:H27), is
RT synthesized via a precursor molecule.
RL Mol. Microbiol. 6:1539-1546(1992).
CC -I- FUNCTION: THIS IS AN ADHESION PROTEIN NECESSARY FOR THE DIFFUSE
CC ADHERENCE OF DIARRHEA-CAUSING ENTEROPATHOGENIC ESCHERICHIA COLI
CC TO EPITHELIAL CELLS.
CC -I- SUBCELLULAR LOCATION: Outer membrane.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-slb.ch/announce/
CC or send an email to license@isb-slb.ch).
CC
DR EMBL: X65022; CAA46156.1; -.
KW Cell adhesion; Signal; Outer membrane; Plasmid.
FT SIGNAL 1 49
FT CHAIN 50 1286 ADHESIN AIDA-I.
FT PROPEP 2 1286
SQ SEQUENCE 1286 AA; 132271 MW; B2A00F72AC05FB34 CRC64;

```

alignment_scores:

```

Quality: 247.50 Length: 1470
Ratio: 0.392 Gaps: 62
Percent Similarity: 42.925 Percent Identity: 18.163

```

alignment_block:

US-09-303-518D-653 x AIDA_ECOLI ..

Align seg 1/1 to: AIDA_ECOLI from: 1 to: 1286


```
506 AspGlyGlnValThrGlyThrAsnIleuysnGlyGlyThrIleAr 612
2178 TACAGCTTCGAGCTGACGGCTGACA..... 2205
612 gValAspSerGlyAlaSerAlaLeuAsnIleAlaLeuSerSerGlyAla 629
2206 .....AGTTGTACCGAAAAACATTACCGAGCATTAAGTATTGCT 2247
629 snLeuPheThrSerThrGlyAlaThrLeuProGluLeuThrThrMetAla 645
2248 TCATTGAGC...AAGACCGACATAGAGCAATGCTACGCTTGCCTATCA 2294
646 AlaLeuSerValSerGlnAsnHisAlaSerAsnIleValLeuGlnAsnI 662
2295 CGCTCATTTAAATCTACA.....GGAGCTGCC.....ACACTCA 2329
662 yGlyLeuLeuArgValThrSerGlyGlyThrAlaThrAspThrThrVala 679
2330 ACGGC.....AATCTAGTACAGCGGAGACAGCACTATACG 2367
679 snSerAlaGlyArgLeuArgIleAspAspGlyGlyThrIleAsnGlyThr 695
2368 GTTACGGCGCAACGCC.....ACCCAAA 2390
696 ThrThrIleAsnAlaAspGlyIleValAlaGlyThrAsnIleGlnAsnAs 712
2391 CGGCACCTGACGCTC..... 2406
712 pGlyAsnPheIleuAsnIleuAlaGlnAsnTyAspPheGluThrGlu 729
2406 ..... 2406
729 euSerGlySerGlyValLeuValLysAspAsnThrGlyIleMetThrTy 745
2407 .....GTGGCAATGCCCAACG 2423
746 AlAGlyThrLeuThrGlnAlaGlnGlyValAsnValLysAsnGlyIle 762
2424 AACATTTAATCAAGCCACATTAAAGCGCAACACATCGGCTTCGGAATG 2473
762 eIlePheAspSerAlaValAlaAsnAlaAspMetAlaValAsnGlnAsn 779
2474 CTTCAATTATTTACGACACAGCCGCTACAAAGCGAGCTGACGCTT 2523
779 IatyrIleAsnIleSerAspGlnAlaThrIleAsnGlySer.....Val 793
2524 TCCGACACGCTAAGCAAAAGTACGATTCGCACTCAAGCGCAATGT 2573
794 AsnAsnAsnGlySerIleValIleAsnAsnSerIleIleAsnGlyAsnI 810
2574 CTCCCTAAGCGGATAGGACAGTATTCATTTGAAAACAGCCGCTTACCG 2623
810 eThr.....AsnAspAlaAspLeuSerPheGlyThrAlaLys..... 822
2624 GAAAAATCAAGCGCGGCAAGATACGCAATTACCTTAAAGACACGCA 2673
822 ..... 822
2674 TGGACGCTCGCTCGGCGACGGAATTAGCAATTTAACTTGACAAACGC 2723
823 .....LeuLeuSerAlaThrValAsnGlySerLeu...ValAsnAsnLy 836
2724 CACCATTAACATCAATTCGCGCTATCGACAGATGCGGCGGCGCAAA 2773
836 sAsnIleIleLeuAsn..... 841
2774 CCGGCGAGTGGCGAGATGGCGCGCGCGCTTCGGCGCGTTCCTATTA 2823
842 .....ProThrLysGlnSerAlaGlyAsnThrLeu 851
2824 TCCGTTACCGCGCAACTTCGGCAGAAATCCGTTTCAACACGCTGACGCT 2873
:::|||||:::
506 AspGlyGlnValThrGlyThrAsnIleuysnGlyGlyThrIleAr 612
2178 TACAGCTTCGAGCTGACGGCTGACA..... 2205
612 gValAspSerGlyAlaSerAlaLeuAsnIleAlaLeuSerSerGlyAla 629
2206 .....AGTTGTACCGAAAAACATTACCGAGCATTAAGTATTGCT 2247
629 snLeuPheThrSerThrGlyAlaThrLeuProGluLeuThrThrMetAla 645
2248 TCATTGAGC...AAGACCGACATAGAGCAATGCTACGCTTGCCTATCA 2294
646 AlaLeuSerValSerGlnAsnHisAlaSerAsnIleValLeuGlnAsnI 662
2295 CGCTCATTTAAATCTACA.....GGAGCTGCC.....ACACTCA 2329
662 yGlyLeuLeuArgValThrSerGlyGlyThrAlaThrAspThrThrVala 679
2330 ACGGC.....AATCTAGTACAGCGGAGACAGCACTATACG 2367
679 snSerAlaGlyArgLeuArgIleAspAspGlyGlyThrIleAsnGlyThr 695
2368 GTTACGGCGCAACGCC.....ACCCAAA 2390
696 ThrThrIleAsnAlaAspGlyIleValAlaGlyThrAsnIleGlnAsnAs 712
2391 CGGCACCTGACGCTC..... 2406
712 pGlyAsnPheIleuAsnIleuAlaGlnAsnTyAspPheGluThrGlu 729
2406 ..... 2406
729 euSerGlySerGlyValLeuValLysAspAsnThrGlyIleMetThrTy 745
2407 .....GTGGCAATGCCCAACG 2423
746 AlAGlyThrLeuThrGlnAlaGlnGlyValAsnValLysAsnGlyIle 762
2424 AACATTTAATCAAGCCACATTAAAGCGCAACACATCGGCTTCGGAATG 2473
762 eIlePheAspSerAlaValAlaAsnAlaAspMetAlaValAsnGlnAsn 779
2474 CTTCAATTATTTACGACACAGCCGCTACAAAGCGAGCTGACGCTT 2523
779 IatyrIleAsnIleSerAspGlnAlaThrIleAsnGlySer.....Val 793
2524 TCCGACACGCTAAGCAAAAGTACGATTCGCACTCAAGCGCAATGT 2573
794 AsnAsnAsnGlySerIleValIleAsnAsnSerIleIleAsnGlyAsnI 810
2574 CTCCCTAAGCGGATAGGACAGTATTCATTTGAAAACAGCCGCTTACCG 2623
810 eThr.....AsnAspAlaAspLeuSerPheGlyThrAlaLys..... 822
2624 GAAAAATCAAGCGCGGCAAGATACGCAATTACCTTAAAGACACGCA 2673
822 ..... 822
2674 TGGACGCTCGCTCGGCGACGGAATTAGCAATTTAACTTGACAAACGC 2723
823 .....LeuLeuSerAlaThrValAsnGlySerLeu...ValAsnAsnLy 836
2724 CACCATTAACATCAATTCGCGCTATCGACAGATGCGGCGGCGCAAA 2773
836 sAsnIleIleLeuAsn..... 841
2774 CCGGCGAGTGGCGAGATGGCGCGCGCGCTTCGGCGCGTTCCTATTA 2823
842 .....ProThrLysGlnSerAlaGlyAsnThrLeu 851
2824 TCCGTTACCGCGCAACTTCGGCAGAAATCCGTTTCAACACGCTGACGCT 2873
:::|||||:::
852 ThrValSerAsnTyThrGlyThrProGly.....SerValIleSerIle 866
2874 AAACGGCAATTAAGACGTCAGGAAACATTCGCTTATGTCGAGACTCT 2923
866 uGlyGlyValLeuGluGlyAspAsnSerLeuThrAspArgLeuValVal 883
2924 TCGGCTACCCGACCGCGCAATTTAAGCTGGCGGAAAGTTCCGAGGACT 2973
883 ySdIysnThrSerGlyGln.....SerAspIleVal 893
2974 TACACCTTGGCTGTCAACAATACCGCAACGCAACCCGTAAGTCTGAGCA 3023
894 Tyr.....ValAsnGlnAspGlySerGlyGlyGlnThrArgAspI 907
3024 ATTGACGGTAGTGAAGAAAGACACACACCGCTGTCCGAAATCTTA 3073
907 yIleAsnIleIleSerValGluGlyAsnSer.....AspAlaG 920
3074 ATTTCACCTGCGCAAAACGAAACACGTCGATCCGCGGCGCATGCGTTATCAG 3123
920 IuPheSerLeuLysAsn...ArgValAlaIleGlyAlaTyAspTyThr 935
3124 CTTATCCGCAAGACGCGAGTTCGCTCATTAATCCGGTCAAGAACCA 3173
936 Leu..... 936
3174 AGAGCTTTCGACAAACTCGCAGAGCGGGAGAAACAGAGCGGCTTGA 3223
936 ..... 936
3224 CGGCAAAACAGGACACACTTCGCGCAAAACACAGCGGAAACACAC 3273
937 ..... 940
3274 GCGCAAAAGCTTGACGCGCTGATTCGCGCGGCGCAATGCCACGAAA 3323
940 ..... 940
3324 GGCAGAAAGTGTGGCGGAACCGCGCGGACGAGCGGGAATAATGCGG 3373
941 .....GlnIysGlyAsn 940
3374 GCATTATGACAGCGGAGGAAGAAAGAAACGGGTGCAGCGGATTAACAC 3423
948 IYTrpTy..... 950
3424 ACCGCTTGGCGAAACAGCGGAAACCGCGCGGCTACACCGC 3473
951 .....LeuThrSerHl 954
3474 CTTCCCGCGCGCGCGCGCGCGCGGATTTGCCGAACCGACGCGCC 3523
954 sLeuProThrSer.....AspThrArgGlnTyArgProG 966
3524 AACCGCAACCCCAACCGGAGCGAGCTGATCAGCCGTTATGCAATAGC 3573
966 Lu.....AsnGly 968
3574 GGTTAGATGAATTTTCGCGACGCTCAACAGCTTTTCGCGTACAGGA 3623
969 SerTyAlaThrAsnMetAlaLeuAlaAsnSerLeuPheLeuMetAspLe 985
3624 CGAATTGGACGCGGCTT.....GCCGAAGACCGCC 3655
985 uAsnGluArgLysGlnPheArgAlaMetSerAspAsnThrGlnProGlu 1002
3656 GCAAGCGCGTTTG.....ACAAGCGGATCCGCGGACACCAACAC 3696
1002 eAlaSerValTrpMetLysIleThrGlyLysSerSerGlyLys... 1017
3697 TACGTTGCGAAGATTTCGCGGCTTACCGCAACCAACGACCTGGCGCA 3746
1017 ..... 1017
```

```

3747 AATGGTATGACAGAAAAACCTCGGACACGGGCGGTCGGCATCTGTTT 3796
      |||:|||||:
1018 .....LeuasnaspGlyIn..... 1022
3797 CGCACAAACGGACCGGAACACCTTCAGACAGGCACTGGCACTCGCA 3846
      |||:|||||:
1023 ..AsnLysThrThrAsnGlnPheLeasnGlnLeuGlyGlyAspIle 1038
      |||:|||||:
3847 CGGCTTGCCGACGGTGGCGGTTTCGGCAATACGCACTCGGAGTTTGA 3896
      |||:|||||:
1039 TyrLysPheHisAlaGlu.....GlnLeuGlyaspPheTh 1050
3897 CATCGGATATCAGCGGGGCGGCTTTTACGCGGACGCTTTCACACG 3946
      |||:|||||:
1050 rLeuGlyIleMetGlyTyrAlaAsnAlaLysGlyLysThrIleasnT 1067
3947 GCATCAGAGGCAAAATCCGCGCGCGGTTCGATTAACGGCATTCAGCA 3996
      |||:|||||:
1067 yrrhrSerAsnLysAlaAlaAsnThrLeuasp..... 1078
3997 AGATACCGCGAGGTTTCGGGATTCGCGATCGACACGACATCGGCGC 4046
      |||:|||||:
1079 GlyTyrSerValGlyValTyrGlyThrTyrGlnAsnGlyGlnAsnAl 1095
4047 AACGGCTATTTCGTCACAAAA..GGCGATTACGATACGAAACGCA 4093
      |||:|||||:
1095 aThrGlyLeuPheAlaGlnThrTyrPmetGlnTyrAsnThrPheAsnAla 1112
4094 ATATCGCACCCCGGCGCTTCGATTCACACGCTACCGC..... 4131
      |||:|||||:
1112 ervallGlyGlyaspGlyLeuGlnGlnGlyTyrAsnLeuAsnGlyLeu 1128
4131 ..... 4131
1129 ThrAlaSerAlaGlyGlyTyrAsnLeuAsnValHisThrTyrPheSe 1145
4132 ...GGCGCATTTAGCGAGATTTATTCATTCACACG.....GGCG 4168
      |||:|||||:
1145 rProGlnGlyIleThrGlnGlyLeuPheThrLeuGlnProHisLeuGlnAla 1162
4169 AACACATTTCCATCAGCGCTTATTTAGCCTCTCTATACGATCCGCT 4218
      |||:|||||:
1162 aLrrPmetGlyValThrProAspThrHis..... 1171
4219 TCCGGCAAGTCCGAACGGCGCTCAATACCGCGTATTCGGCAGATTT 4268
      |||:|||||:
1172 .....GlnGlnAspAsnGlnGlyThrValGlnGlyAl 1182
4269 CGGCAAAACCCGC.....AGTGGCAATGGGCGC 4297
      |||:|||||:
1182 agLysAsnAsnIleGlnThrLysAlaGlyLeuArgAlaSerTyrLysV 1199
4298 TAAACGCCCAATCAAAAGTTTCACGCTGCTCCACCGCTCCCGCGC 4347
      |||:|||||:
1199 aLlySer.....ThrLeuAspLysThrPheThrArg 1210
4348 AAGGGCGCGCAATTCGACGCGAG.....CACAGGCGCG 4382
      |||:|||||:
1211 pheArgProTyrIleGlnAlaAsnThrPheHisAsnThrHisGlnPhe 1227
4383 CATCAATTA 4392
      |||:|||||:
1227 yValLysMet 1230

```

seq_name: SwissProt_40:BIGA_SALTY

seq_documentation_block:

ID BIGA_SALTY STANDARD; PRT; 1953 AA.
AC P25927; P25928; O9XC03;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAR-2002 (Rel. 41, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)

```

DE putative surface-exposed virulence protein bigA precursor.
GN BIGA OR STM3478.
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; gamma subphylum; Enterobacteriaceae;
OC Salmonella.
CX NCBI_TaxID=602;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=14028.
RA Stojiljkovic I., Valentine P., Heffron F.:
RT "Salmonella typhimurium rns homolog.":
RL submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
[2]
RP SEQUENCE FROM N.A.
RC STRAIN=LT2 / SGSC1412 / ATCC 700720;
RX MEDLINE=21534948; PubMed=11677609;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA Courtney L., Portwollik S., All J., Dante M., Du F., Hou S., Layman D.,
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA Waterston R., Wilson R.K.:
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
RT LT2.":
RL Nature 413:852-856(2001).
[3]
RP SEQUENCE OF 1-765 FROM N.A.
RC STRAIN=LT2.
RX MEDLINE=91100301; PubMed=1987123;
RA Wu J.Y., Siegel L.M., Kredich N.M.:
RT "High-level expression of Escherichia coli NADPH-sulfite reductase:
RT requirement for a cloned cysG plasmid to overcome limiting sigmaHeme
RT cofactor.":
RL J. Bacteriol. 173:325-333(1991).
CC -!- CAUTION: Ref.3 sequence differs from that shown due to frameshifts
CC in positions 414 and 732.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; AF133696; AAD39458.1; -.
DR EMBL; AE008859; AAL22340.1; -.
DR EMBL; M64606; AAA27042.1; ALT_FRAME.
DR EMBL; M64606; AAA27043.1; ALT_FRAME.
DR PIR; C39200; C39200.
DR PIR; D39200; D39200.
DR StGene; SG10437; bigA.
DR Virulence; Repeat; Signal; Complete proteome.
KW CHAIN
FT FT 1 27
FT FT 28 1953
FT FT
FT DOMAIN 101 252
FT REPEAT 101 103
FT REPEAT 104 113
FT REPEAT 114 122
FT REPEAT 123 133
FT REPEAT 134 144
FT REPEAT 145 155
FT REPEAT 156 166
FT REPEAT 167 177
FT REPEAT 178 188
FT REPEAT 189 199
FT REPEAT 200 210
FT REPEAT 211 221
FT REPEAT 222 232
FT REPEAT 233 243
FT REPEAT 244 252
FT REPEAT 207 252
FT CONFLICT 514 514
FT CONFLICT 514 514

```


FT CONFLICT 1698 1698 D -> N (IN REF. 1).
 FT CONFLICT 1795 1798 OYLE -> TLO (IN REF. 1).
 FT CONFLICT 1836 1837 SA -> T (IN REF. 1).
 SO SEQUENCE 1953 AA; 200150 MW; 611B5F1C954D91AE CRC64;

alignment_scores:

Quality: 245.00 Length: 1687
 Ratio: 0.314 Gaps: 80
 Percent Similarity: 46.236 Percent Identity: 19.265

alignment_block:

US-09-303-518D-653 x BIGA_SALTY ..

Align seg 1/1 to: BIGA_SALTY from: 1 to: 1953

```

217 GTTACACAAAAAGGGAGTGGTCGCAATCCATGACGAAAGCCC 266
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
413 ValAspAsnAlaThrGlyAlaLeuIleSerGlyAsnGlyThrThrAs 429
    ::::: ::::: ::::: ::::: ::::: ::::: :::::
267 GATGATGATTTTCTGTCGTCATCGCGTACGCGTGGCGCATTTGGCG 315
    ::::: ::::: ::::: ::::: ::::: ::::: :::::
429 nPheAlaGlyAspIleAlaValSerGlyGlyGlyThrAlaIleIleA 446
    ::::: ::::: ::::: ::::: ::::: ::::: :::::
316 .GGCGATCAATATTATTGTCAGCGTGCACATACGCGCTTAACAA 363
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
446 spGlyAspAsnAlaThrIleLys.....AsnThrGlyThrSerAsp 459
    ::::: ::::: ::::: ::::: ::::: ::::: :::::
364 GTTGAT...TTGTGGGAGGAGGAGCAATCCCGATCAGCAGCGCTTTC 410
    ::::: ::::: ::::: ::::: ::::: ::::: :::::
460 IleSerGlyAlaGlySerThrGlyThr..... 468
    ::::: ::::: ::::: ::::: ::::: ::::: :::::
411 TTACCAATTGTGAAAAAATATTATTAA..... 441
    ::::: ::::: ::::: ::::: ::::: ::::: :::::
469 .....ValIleAspGlyAsnAsnAlaArgValAsnAsnAspGlyAspM 483
    ::::: ::::: ::::: ::::: ::::: ::::: :::::
442 .....GCAGGACTAAGCGGCTATGCGCGCGCTTAT 477
    ::::: ::::: ::::: ::::: ::::: ::::: :::::
483 ethrIleThrAspGlyGlyThrGlyGlyIleThrGlyAspAsnVal 499
    ::::: ::::: ::::: ::::: ::::: ::::: :::::
478 CATATGCGCGTTTGACAAATTTGTCACAGATCGAGACCTGTGAGAT 527
    ::::: ::::: ::::: ::::: ::::: ::::: :::::
500 ValIleAspAsnAlaGlySer.....ThrIleValSerGlyAlaAspAl 514
    ::::: ::::: ::::: ::::: ::::: ::::: :::::
528 GACCACT...TATATGATGGGTGGAATACGCTGATTTAATAATACC 574
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
514 athrAlaLeuTyrlleGlyIleGlyIleAspAsnAlaLeuValIleAsnGlyGlyA 531
    ::::: ::::: ::::: ::::: ::::: ::::: :::::
575 CTGATCTGTTCGAATCGAGCA...GGCAGACAAATATTGGCGGCTCAT 621
    ::::: ::::: ::::: ::::: ::::: ::::: :::::
531 snGlnThrIleSerGlyGlyAlaValGlyThrGlyIleAspGlyAsp 547
    ::::: ::::: ::::: ::::: ::::: ::::: :::::
622 GAAGACGAACCAATACCGCGAAGTTTCATATTGTCAGAGCGCAT 671
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
548 AlaIleThrThrAsnThrGlyAspIleAlaValAspGlyAlaGlySerAl 564
    ::::: ::::: ::::: ::::: ::::: ::::: :::::
672 TTCTTGGCTCGTGGGCGCAATAC..... 696
    ::::: ::::: ::::: ::::: ::::: ::::: :::::
564 alaValIleIleAsnGlyAspAsnGlySerLeuThrGlnAlaGlyAspL 581
    ::::: ::::: ::::: ::::: ::::: ::::: :::::
697 .....TTGCACAAATGATGATCAGCTGT 720
    ::::: ::::: ::::: ::::: ::::: ::::: :::::
581 eulLeuValThrAspGlyAlaMetGlyIleIleThrTyrlleGlyThrGlyAsn 597
    ::::: ::::: ::::: ::::: ::::: ::::: :::::
721 GGCACAGCACTTAGTACGCAAAAAATTAACATAC...CCATATGG 767
    ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
598 GluAlaValAsnThrGlyAsnAlaThrValArgAspAlaAspSerValG 614
    ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
768 TTTTTCACACAGCA.....GGCTCATTT.....GGCGAC... 798
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
614 yPheValAlaGlyGlyIleAsnThrPheLysAsnGlySerAlaValH 631
    ::::: ::::: ::::: ::::: ::::: ::::: :::::
799 .....AGTGGCTCACCAATGTTTATCTATGATGCCCAAAAGCA 837
    ::::: ::::: ::::: ::::: ::::: ::::: :::::

```

```

631 spValSerLeuAsnGlyThrGlyAlaLeuValSerGlyAspMetSerGln 647
    ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
838 AACTGGTTA.....ATTAAT.....CG 854
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
648 ValThrLeuAspGlyAspIleAsnValAlaSerValGlnAspSerGluG 664
    ::::: ::::: ::::: ::::: ::::: ::::: :::::
855 GGTATTGCAA.....ACAGCAACCCCTATATAGCAAAAAACCAATGGCT 898
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
664 yAlaPheSerSerAlaThrGlyValSerValSerGlyAspSerAlaVal 681
    ::::: ::::: ::::: ::::: ::::: ::::: :::::
899 TCACGCTA.....GTTCTAAAGATTGGTTCTATATAGTAA 933
    ::::: ::::: ::::: ::::: ::::: ::::: :::::
681 AlaSpIleThrGlyAsnValAlaSnIleSerAlaAspTyrlleGlnAsp 697
    ::::: ::::: ::::: ::::: ::::: ::::: :::::
934 ATCTTGTGCTGAGATACCATTCAGATTTCTACGAAACCAATCAAAATGG 983
    ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
698 LeuAlaAlaGly.....AlaProPoleuThrGly 707
    ::::: ::::: ::::: ::::: ::::: ::::: :::::
984 GAAATACCTTTTAAACGACATTAAT.....ATGCGCGAGGAA 1021
    ::::: ::::: ::::: ::::: ::::: ::::: :::::
707 yValValValGlyGlyAsnGlyAsnThrValThrLeuAsnGlyAlaLeuA 724
    ::::: ::::: ::::: ::::: ::::: ::::: :::::
1022 AAATGATGCCCAACATAACACTTTCTCTACCTTATAGTTAAACA 1071
    ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
724 snIleAspAsnAsnSer.....AlaThr 733
    ::::: ::::: ::::: ::::: ::::: ::::: :::::
1072 CGAACCGTTCATTTGTTAATGTTTCTTATCCGACAGCAAGAGAAC 1121
    ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
734 GlyGlyGlnTyrlleuAspValAlaGlyLeuSerValThrGlyAspAsp 750
    ::::: ::::: ::::: ::::: ::::: ::::: :::::
1122 TGTTATATGCTGCGAGTGGGCTCAC.....AGTTATGACCCA 1162
    ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
750 nAspValGluIleAspGlyGlyIleAsnIleThrHisSerGluAspProl 767
    ::::: ::::: ::::: ::::: ::::: ::::: :::::
1163 GACTCAATATGAGAGAAATATTCCTTATTGAC...AAGAGAAAGGT 1209
    ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
767 euAspGlyThrSerAlaAspIleThrGlyIleSerValSerGlyAsnSer 783
    ::::: ::::: ::::: ::::: ::::: ::::: :::::
1210 GAATGATACTT.....ACGACACATCAACCAAGCGCGG 1247
    ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
784 ThrValThrLeuAsnGlyHisSerThrIleAspThrAsnThrValAlaG 800
    ::::: ::::: ::::: ::::: ::::: ::::: :::::
1248 CGGTTTGATTTGAGGTATTTAGCGTCTGCCCAAAACACAGCAA 1297
    ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
800 Y.....GlyHisValValLeuAlaArgValAsnAsn.... 810
    ::::: ::::: ::::: ::::: ::::: ::::: :::::
1298 CGTGGCAAGCGCGCGGCTTCATATCATGATGTCAGTACCTGTTACTTG 1347
    ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
811 .....GlyGlySerLeuIleGluGlyAspAspSerValAlaAspVal 824
    ::::: ::::: ::::: ::::: ::::: ::::: :::::
1348 AAAGTAAACGGCGTGGCA.....AACGACCGCTGTC 1379
    ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
825 AsnValSerTyrlleProThrGlyTyrlleThrTyrlleAsnAlaLeuMet 841
    ::::: ::::: ::::: ::::: ::::: ::::: :::::
1380 CAAATCGGCAAGCAAGCAGCTGCTGTTCAAGCAAGAGGAA..... 1422
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
841 tAlaAspGlyGluGlyThr...SerIleGluAsnGlyGlyAspIleHis 857
    ::::: ::::: ::::: ::::: ::::: ::::: :::::
1423 .....AACCAAGCTCGTCAAGCTG 1443
    ::::: ::::: ::::: ::::: ::::: ::::: :::::
857 erHisGlyValTyrlleSerValIleArgAlaAspAsnGlySerGluValSer 873
    ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1444 GGCACGCTAAAGCATCTTAGATACACAGCGAGCATCAAGCAAAAA 1493
    ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
874 AsnSerGlyAspIleLeuValTyrlleAlaThrSerIleAsnSerGluAs 890
    ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1494 ACAAGCTTTAGTGAATCGCTTGGTCAGCGGCGAGGCGAGCTGCAAC 1543
    ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
890 pArgAla.....AlaIleThrArgAlaSerGlyGlyGlySerAlaValH 905
    ::::: ::::: ::::: ::::: ::::: ::::: :::::
1544 TGAAT.....GCGGATATACAGTCAACCCGACCAAACTCTATTTC 1584
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```



```

3112 TGGCGTTATCAGCTTATCCGCAAGACGGAGTCCGCTG..... 3153
      :::: ||| :::: ||| :::: ||| :::: |||
1451 GlyThrThrLysAlaIleIleLeuSngIuIleAsnLeuLeuCysAs 1467
3154 ..... 3154
1467 pSerGlyCysAspIleTyrAlaProGlyThrThrGlyThrGlnAspH 1484
3155 ATAAATCCGGTCAAGACAGAGCTTCCGCAAACTCGGCAAGCGGGA 3204
      |||| :::: ||| :::: ||| :::: |||
1484 IAsnGlyThrAlaAspIleValIleProAsp.....AlaThrAla 1498
3205 GAACAGAGCGCGCTTACGCGCAAAACAGCAACTGCGCGCAACA 3254
      |||| :::: ||| :::: ||| :::: |||
1499 ProThrGlnGlySerIleProThrProProAlaAspProAlaProG 1515
3255 ACAGCGGGA..... 3267
      |||| :::: ||| :::: ||| :::: |||
1515 nGlnLeuSerAsnTyrIleValGlyThrAsnAlaAspGlySerSerG 1532
3268 ..... 3288
      |||| :::: ||| :::: ||| :::: |||
1532 hrLeuLysAlaAsnAsnLeuValIleGlyAspAsnValLysValAsp 1548
3289 GCGCTGATTGCGCGCGCGCAATGCCGCAAGAGCAAGTGTTC 3338
      :::: ||| :::: ||| :::: ||| :::: |||
1549 GlyPheThrSerGlyThrAlaAspThrThrValValAlaAspAlaPh 1565
3339 GCAACGCGCGCGCGAGCGCGGGAAT.....GCCGCGATTA 3379
      :::: ||| :::: ||| :::: ||| :::: |||
1565 eThrGlySerAsnIleGlnGlyAlaAspAsnIleThrSerThrSerVal 1582
3380 TCCAGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3429
      :::: ||| :::: ||| :::: ||| :::: |||
1582 alTyrAsnAlaGlnGlySerGlnAspAlaAspGlyAsnValAspAlaThr 1598
3430 TTGGCGGAACAGCGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3479
      :::: ||| :::: ||| :::: ||| :::: |||
1599 MetThrLysAsnAlaTyrAlaAspValAlaThrAspSerSerVal..... 1613
3480 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3529
      |||| :::: ||| :::: ||| :::: |||
1614 .....SerAspValAlaGlnAla..... 1619
3530 AACCCCAACCGCAGCGCGATGATCAGCGCTTATGCCAATAGCGGTTG 3579
      ||| :::: ||| :::: ||| :::: |||
1620 .....LeuAspAlaGlyTyrThrAsnAsnIleLeu 1629
3580 .....AGTAATTTCCGCCAGCGTCAA 3602
      :::: ||| :::: ||| :::: |||
1630 TyrThrSerLeuAsnValGlyThrThrAlaGlnLeuAsnSerAlaLeu 1646
3603 CAGCGTTTGGCGGTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3652
      :::: ||| :::: ||| :::: |||
1646 GlnValSerGlyAlaGln.....AlaThrThrValPheArgIuAla 1661
3653 GC..... 3654
      ||| :::: ||| :::: ||| :::: |||
1661 rgValLeuSerAsnArgPheThrMetLeuAlaAspAlaAlaProGlnIle 1677
3655 .....CGCAAGCGCGTTGGCAAGCGGATCCG..... 3684
      |||| :::: ||| :::: ||| :::: |||
1678 LysAspGlyLeuAlaPheAsnValValAlaLysGlyAspProArgAlaG 1694
3685 .....GACACCAACACTACCGTTCCAGAGATTTCGCGCTAAC 3724
      |||| :::: ||| :::: ||| :::: |||
1694 uLeuGlyAsnAspThrGlnTyr.....AspMetLeuAlaLeu 1707
3725 GCCAACAACCGACCTCGC.....CAATCGGT 3753
      |||| :::: ||| :::: ||| :::: |||
1707 rgGlnThrLeuAspLeuThrAlaSerGlnAsnLeuThrLeuGlnTyrGly 1723
3754 ATGCAGAAACCTCGGAGCGCGCGTCCGATCTCTGTTTGCACAA 3803

```

```

      :::: ||| :::: ||| :::: |||
1724 IleAlaTyrLeuAspGlyAspGly.....SerLysH 1734
3804 CCGGACCGGAAACACCTTCGACGACGACGACGACGACGACGACGACG 3853
      ||| :::: ||| :::: ||| :::: |||
1734 rAlaGlyAspAsnGlyLeuThrGlyGlyTyrSerGlnPhePheGlyLeu 1751
3854 CCCAGCGTCCGCTTTTGGCGAATACGCAATCGGC..... 3888
      |||| :::: ||| :::: ||| :::: |||
1751 ySHisSerMetAlaPheAspGlu.....GlyLeuAlaTyrAsnAsnSerLeu 1766
3889 AGTTTCGACATC...GGCATCAGCGCGCGCGCGGCTTTAGTAGCGGAC 3935
      |||| :::: ||| :::: ||| :::: |||
1767 ArgTyrAspValHisAsnLeuAspSerSerArgSerValAlaTyrGlyAs 1783
3936 CCTTCAGACGCGCATCAGAGCAAAATCCGCGCGCGCTGCATATAC. 3984
      :::: ||| :::: ||| :::: |||
1783 pValAsnLysIleAlaAspSerAspMetArgGlnTyrLeuGlnPheA 1800
3985 .....GGCATTCAGCAAGATACCGCGAGTTTCGCGCGATTCGGC 4026
      |||| :::: ||| :::: ||| :::: |||
1800 rgSerGlnGlyAlaLysThrPheThrMetMetGly...AspAlaLeuLys 1815
4027 ATCGAACCGCATCGCGCGACGCGCTATTGTCGCAAAAGCGGATTA 4076
      :::: ||| :::: ||| :::: |||
1816 ValThrProTyrAlaGlyValLys.....PheArgHisThrMetGluAs 1830
4077 CCGATACGAAACGTCATATCGCCACCGCGCGCTTGATTCATACCGCT 4126
      |||| :::: ||| :::: ||| :::: |||
1830 pGlyTyrLysGlnArgSerAlaGlyAspPheAsnLeuSerMetLsnSerG 1847
4127 ACCGC.....GCCGCGATTAAAGCAGATTTATCA 4155
      :::: ||| :::: ||| :::: |||
1847 LysGlnGlyThrAlaValAlaAspSerIleValGlyLeuLysLeuAspTyrAla 1863
4156 TTCAACCGCGCGCAACACATTTCCATCACG.....CCTTATT 4193
      ||| :::: ||| :::: ||| :::: |||
1864 GlyLysAspGlyTyrPserAlaThrAlaThrLeuGlnGlyLysProAsnIle 1880
4194 GAGCTTCCTATACGATGCCGCTTCGCGCAAAAGTCGACGCGGCTCA 4243
      |||| :::: ||| :::: ||| :::: |||
1880 uSerTyrSerLysSerGln.....ArgThrAlaSerL 1891
4244 ATACCGCGCTATTGCGCGAGATTTGCGCAAAACCGGAGTGGGAGATG 4293
      :::: ||| :::: ||| :::: |||
1891 euGlnGlyAlaAlaGlyGlnSerPheGlyValAspAspGlyGlnLysGly 1907
4294 GCGCTAAACCGCGAAATCAAGGTTTCACGCTGCCCTCCACGCTCGCG 4343
      ||| :::: ||| :::: ||| :::: |||
1908 Gly.....GlyValAsnGlyLeuAl 1914
4344 CGCCAGGCGCGCAATTTGAGACGCGACGACGCGGCGGATGAATTAG 4393
      |||| :::: ||| :::: ||| :::: |||
1914 aThrIleGlyValLysTyrSerSerAsnAspThrAlaLeuHisLeuAspA 1931
4394 GCTACCGCTGG 4404
      ::||| :::: |||
1931 latYrGlnTyr 1934

seq_name: SwissProt_40:YEJO_ECOLI
seq_documentation_block:
ID YEJO_ECOLI STANDARD: PRT: 863 AA.
AC P33924: P76450: Q47291:
DT 01-FEB-1994 (rel. 28, created)
DT 01-FEB-1994 (rel. 28, last sequence update)
DT 16-OCT-2001 (rel. 40, last annotation update)
DE Hypothetical outer membrane protein yefJ.
GN YEJO OR B2190.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;

```

```

RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / BHB2600;
RA Richerich F., Lakey N., Gryan G., Jaehn L., Mintz L., Robison K.,
RA Church G.M.;
RL Submitted (OCT-1993) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RX MEDLINE-97426617; PubMed-9278503;
RA Blatner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Colado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [3]
RP PRESENCE OF AN INSERTION SEQUENCE.
RA Rudd K.E.;
RL Unpublished observations (JAN-1994).
CC -1- SUBCELLULAR LOCATION: Outer membrane (Potential).
CC -1- SIMILARITY: STRONG, TO BORDETELLA PERTACTIN.
CC -1- CAUTION: THIS IS A CONCEPTUAL TRANSLATION; THE READING FRAME IS
CC INTERRUPTED BETWEEN CODONS 21 AND 22 BY A 155K INSERTION ELEMENT.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U00008; AAA16385.1; ALT_SEQ.
DR EMBL; AE000308; AAC75250.1; ALT_SEQ.
DR EcoGene; EGI2051; yeJ0.
KW Hypothetical protein; Outer membrane; Complete proteome.
SQ SEQUENCE 863 AA; 91202 MW; 55D460B5BC3D9AD CRC64;

alignment_scores:
Quality: 243.00 Length: 1130
Ratio: 0.488 Gaps: 51
Percent Similarity: 44.071 Percent Identity: 19.646

alignment_block:
US-09-303-518D-653 x YEJO_ECOLI ..

Align seg 1/1 to: YEJO_ECOLI from: 1 to: 863

1213 TTGATACCTTACCAGACATCAACCAAGC.....GCGGGCGTTTGTGA 1256
:: :|||:::||||::: ||| ::|::: |||:::
26 lIeAlaLeuAlaSerThrValGluTrpGluThrValAspGlyValVal 42
:: :|||:::||||::: ||| ::|::: |||:::
1257 TTTTGAGGGTAAATTTTACGGTC.....TCGCTTAAAAACAAGAAA 1297
:: :|||:::||||::: ||| ::|::: |||:::
42 lIeuGluLysAspIleGlnLeuValTyrGlyThrAlaAsnSerThrLys 59
:: :|||:::||||::: ||| ::|::: |||:::
1298 CGTGGCAAGCGCGGGCGCTTCAATACGATGACGATACCGTACTGG 1347
|||::: |||::: |||::: |||:::
59 lIeAsnProGlyGlyGlnGlnHisIle..... 67
1348 AAGATTAAGCGCGTGGCAACAGCAGCGCTGTCCAAATTCGCAAGGAC 1397
|||::: |||::: |||::: |||::: |||::: |||:::
68 LysGluPheGlyValSerAsn.....ThrGluIleAsnGlyGly 82
1398 GCTGCTGTTCAAGCCAAAGGGGAAACCAAGGCTCGTACGCGTGGCG 1447
:: :|||::: |||::: |||::: |||::: |||::: |||:::
82 rGlnTyrIleGluMetAsnGlyAlaAlaGluTyrSerVal...LeuAsn 98
1448 ACGGTAAATCACTTAA.....GATCAGCAGCGCGAC 1479
|||::: |||::: |||::: |||::: |||::: |||:::
98 sPcIlyrGlnIleValGlnMetGlyAlaAlaAlaSerIleThrLeu 114

```

```

1480 GATCAAGGCAAAAAACAAGCCTT.....AGTAAATGGCTTGGT 1520
::: |||::: |||::: |||::: |||::: |||::: |||:::
115 AsnAsnGlyValLeuGlnValTyrGlyAlaAlaAsnAspThrThrLeu 131
1521 CAGCGCGAGGGGAGCGTGCACATGTAATCCGATTAATCAATCCACCCG 1570
::: |||::: |||::: |||::: |||::: |||::: |||:::
131 sGlyGlyArgLeuIleValGluLysAspGlyGlyAlaValPheValAla 148
1571 ACAACATCTTTCCGGCTTCCGGCGGAGCGTTGGATTTGAACGGGCAT 1620
::: |||::: |||::: |||::: |||::: |||::: |||:::
148 lGlu.....LysGlyGlyLeuLeuGlnValTyrGlyGly 159
1621 TCCCTTTCCTTCCACCGCATTTCAAAATACGATGAAGGGCGATGATTGT 1670
::: |||::: |||::: |||::: |||::: |||::: |||:::
160 GlyPheAlaPhe...AlaValAspGlnLysAlaGlyAlaIle..... 173
1671 CAACCAATCAAGACAAAGATCCACCGTTACCATTAAGCAATTAAG 1720
|||::: |||::: |||::: |||::: |||::: |||:::
174 .....LysThrThrThrArgAlaMetG 181
1721 ATATTACTACACCGCAATTAACACACTTGGATAGCAAAAGAAATT 1770
::: |||::: |||::: |||::: |||::: |||::: |||:::
181 lValPheGlyThrAsnArgLeuGlyGlnPheAspIleLysAsnGlyLe 197
1771 GCTTACACCGCTTGTTGGCGAGAAAGATCAACCAACGAGCGCG 1820
|||::: |||::: |||::: |||::: |||::: |||:::
198 AlaAsnAsn..... 200
1821 GCTCAATCTGAATTAACCAACGAGAGCGGATCGCACTTACTGCTTT 1870
::: |||::: |||::: |||::: |||::: |||::: |||:::
201 .....MetLeuLeuG 204
1871 CCGGCGGAGCAAAATTAACGCGAATATCAAGCAAAACGCAAACTG 1920
::: |||::: |||::: |||::: |||::: |||::: |||:::
204 lLysnGlyGlySerLeuArg.....ValGluLysn..... 214
1921 TTTTTCAGCGCAGACCGACCGACCGCTTACATCAATTTAGACCGG 1970
::: |||::: |||::: |||::: |||::: |||::: |||:::
215 .....AspPheAlaTyrAsnThrThrValAspSe 224
1971 GTGCTCAAAAATGAGAGTATCCACACAGGAAATCGTGGGACAGC 2020
::: |||::: |||::: |||::: |||::: |||::: |||:::
224 rGlyGlyLeuLeuGlnValMetAspGlyGlyThrValThr..... 237
2021 ATTGATCGACCGCACATTTAAAGCGGAAACTTCATTTACGGCGGA 2070
::: |||::: |||::: |||::: |||::: |||::: |||:::
238 ..GlyValAspLys.....LysAla.....GlyGly 245
2071 CAGCGGTGTTTCCGCAATGTTGCCAAATGAGAGCGCATTTGGCATTT 2120
::: |||::: |||::: |||::: |||::: |||::: |||:::
246 LysLeuIleValSerThrAsnAlaLeuGlnValSerGlyPro..... 259
2121 AAGCAATCAGCGCCCAAGCAGTTTGGGTGCGCAGCGCATCAACGACCA 2170
::: |||::: |||::: |||::: |||::: |||::: |||:::
260 .....AsnSerLysGlyGlnPheSerIleLysAspGlyValSerLys 214
2171 CAATCTGTACAGTTCCGACTGACGGGTCTGCAAACTTTACCGAAAA 2220
::: |||::: |||::: |||::: |||::: |||::: |||:::
274 sNfYr...GluLeuAspAspGlySerGlyLeuIle..... 284
2221 ACCATTCACGACATTAAGTATGTTCTTCATTAAGCAGACGACATCAG 2270
::: |||::: |||::: |||::: |||::: |||::: |||:::
285 ValMetGlnAspThrGlnAlaIleAspThrIle..... 295
2271 AGCAATGTCAAGCTTCCGATCAGCGTCAATTAATCTCAGAGACTTG 2320
|||::: |||::: |||::: |||::: |||::: |||:::
296 .....LeuAspLysHisAlaThrMetGln..... 303
2321 CCACACTCAACGCGCAATCTTAGTCAGCGGAGACAGCAGCATTAACGTT 2370
|||::: |||::: |||::: |||::: |||::: |||:::
304 .....SerLeuGlyLysAspThrGlyThrLysVal 313

```

```
2371 ACGGCAAGCCAC.....CAAAAGCAACCT 2399
      |||||
314 GluAlaAsnIValIThrAspLeuGlyArgSerTyrGlnAsnIleSerI 330
      |||||
2400 CAGCGCTGTGGCAATGCCCAACCAATTAATCAAGCCACATTAAGC 2449
      |||||
330 eTrpTyrSerSerLysAla...IleSerGluAsnMetValIleAsnAsnG 346
      |||||
2450 GCAACACATCGGCTCGACATCGCTTCAATTAATCAATGCAACACGCC 2499
      |||||
346 LysArgAlaAsnValITrpAlaGlyIThrMetValAsnValSerValArgGly 362
      |||||
2500 GTACAAAGCGGAGCTGACGCTTCCGACACGCTAAGGCAACCTAAG 2549
      |||||
363 .....AsnAspGlyIleLeuGluValMetLysProGlnIleAsnTyrAl 377
      |||||
2550 CCATTCGGCATCAAGGCAATGCTCCAGCCGATTAAGCGCATTTCC 2599
      |||||
377 aProAlaMetLeuValIglYlValValSerGluGlyAlaSerPhe. 393
      |||||
2600 ATTGTGAAACAGCGCTTACCGGAAATACGCGGCAAGATAG 2649
      |||||
394 .....ArgThrHisGlyAlaValAspThr 401
      |||||
2650 GCA.....TTACACTTAAGACAGCAATGAGCGCTGCGCGG 2690
      |||||
402 SerLysAlaAspValSerLeuGluAsnSerValITrpThrIleIleAla 418
      |||||
2691 C.....ACGGAATTAGGCATTTTAAC 2713
      |||||
418 pIleThrIThrIThrAsnGlnAsnIThrLeuLeuAsnLeuAlaAsnLeuAla 435
      |||||
2714 TTGACACAGCCACCATTAACACTCAATTCGCGCTATGACAGATGGCGCA 2763
      |||||
435 eSerAspAlaAsnValIleMet..... 442
      |||||
2764 GGGGCGCAACCGGCGAGTGGCGAGATGGCGCGCGCGCGCGCG 2813
      |||||
443 .....MetAspGluProValIThrArgSer..... 450
      |||||
2814 TTCCCTATTATCCGTTACGCGGCAACCTTGGCAGAAATCCGTTCAACA 2863
      |||||
451 .....SerValIThr.....AlaSerAlaGluAsn...PheIleT 461
      |||||
2864 CGCTGACGTTAAAGGCAATTGAACGGTCAAGGAACATTCGCTTTATG 2913
      |||||
461 hTrpLeuThrIThrAsn...ThrLeuSerGlyAsnGlyAsnIThrIThrIle 476
      |||||
2914 TCGGAACCTGTGGCGTACCGGCGGCAAAATGAAGCTGGCGGAATTC 2963
      |||||
477 ThrAspMetAlaAsnHisGlnSerAspGlnLeuAsnValIThrGlyGlnAl 493
      |||||
2964 CGAAGCACTTACACCTTGGCTGCAACATACGCGCAACGCAACCGGTA 3013
      |||||
493 aThrGlyAspPheLysIlePheValIThrAspThrGlyAlaSerProAla 510
      |||||
3014 GTCTGAGCAATTGAACGGTAGTGAAGAAAGACACACACCGCTGTCC 3063
      |||||
510 lAgIlyAspSerLeuIThrLeuVal.....ThrIThrGlyGly 521
      |||||
3064 GAAATCTTAATTTCACCTCGCAAAACGACAC.....GTGATCGCGG 3107
      |||||
522 GlyAspAlaAlaIlePheIThrLeuGlyAsnAlaGlyValValAspIleG 538
      |||||
3108 CGCATGCGCTTATCAGCTTATCCGCAAGAGCGGAGTTCCGCTGCATA 3157
      |||||
538 yThrITyrGluITyrIThrLeu..... 545
      |||||
3158 ATCCGCTCAAGACACAGCTTTCCGACAACTCGGACGCGGAGAA 3207
      |||||
545 ..... 545
      |||||
3208 ACAGAGCGCGCTTACGCGCAAAACAGCACACTTCCCGCCCAACACA 3257
      |||||

545 ..... 545
      |||||
3258 GCGGAAAAAGACAGCGGCAAAAGCCTTGACGCGCTGATGCGGCGGCG 3307
      |||||
546 .....AspAsn.....GlyA 549
      |||||
3308 GCATGCGCACGAAAAAGCAGAAAGTGTGCCGACCGCGCGGAGCA 3357
      |||||
549 snHisSerITrpSerIThrLeuAlaGluAsn..... 557
      |||||
3358 GCGGCGGAAAAATGCCGCAATTATGACAGCGGAGAGAAAAAGAGGT 3407
      |||||
557 ..... 557
      |||||
3408 GCAGCGGATTAAGACACCGCTTGGCGAAACAGCGGCAACGCAACCC 3457
      |||||
558 .....ArgAlaGlnIleT 562
      |||||
3458 GCGCGGCTACGACCGCTTCCCGCGCGCGCGCGCGCGCGGATTTG 3507
      |||||
562 hTrpSerIThrIThr..... 566
      |||||
3508 CCGCAACCGGACGCCCAACCGCAACCCACGAGCGGACCTGATCAG 3557
      |||||
567 .....AspValIleAs 570
      |||||
3558 CCGTTATGCCAATAGCGGTTGAGTGAATTTCCGCGCACGCTCAACAGC 3607
      |||||
570 mMetAlaAlaIleGlnProLeu...ValIThrAspAlaGluLeuAspIThr 585
      |||||
3608 TTTTGGCGCTACAGGCAATTTGACCGCGCTGTTGCCGAAACCGCGCG 3657
      |||||
586 .....ValArgGluIThrGlyLeuGlySerValLysGlyValSerTyrAsp 599
      |||||
3658 AACGCGCTTTGACAGAGCGGATCCGGGACACAAACACTACCTGTCGA 3707
      |||||
600 ThrAlaMetITrpSerSerAlaIle...AsnIThrArgAsnAsnValIThrITh 615
      |||||
3708 AGATTTCCGCGCTTACCGCCCAACCAACC..... 3735
      |||||
615 rAspAlaGlyValAlaGlyPheGluGlnIThrLeuIThrGlyLeuIThrLeuGlyI 632
      |||||
3736 ..GACCTGCGGCAATCGGTATGCAAAAAACCTCGGCGAGCGGCGGCTC 3783
      |||||
632 lAspSerIThr.....PheSerITrgIuGluSerIThrITeArg 645
      |||||
3784 GGCATCTGTTTTCGACAAACCGGACGGAACCACTTGCAGACGAGCAT 3833
      |||||
646 GlyLeuIlePheGlyITyrSerHisSerIThrIleGlyPheAspITrgIy.. 661
      |||||
3834 CGGCACTCGGACGCGCTTCCGCGGCTGCGCTTTTGGCAATACGCGCA 3883
      |||||
662 .....GlyLysGlyAsnI 666
      |||||
3884 TCGCAGGTTTCGACATCGCATCAGCGGCGGCGGCTTTAGAGCGGC 3933
      |||||
666 lAspSerITyrIThrLeuGlyAlaITyrAlaGlyITrgIuHisGlnAsnGly 682
      |||||
3934 AGCCTTTCAGACGCG.....ATCAG 3953
      |||||
683 AlatyITrAlaITrgIyValITrValITrValITrAspITrgPheITrAlaITrITrLeuITr 699
      |||||
3954 AGGCAAAATCGCGCGCGCTGCTGCATTCGCG..... 3987
      |||||
699 sGlyLysMetSerITrAsnGlyIThrAlaIThrAlaIThrITrAspITrgITrAsnSerITrAsnG 716
      |||||
3988 .....ATTCAGGCAAGATACCGCGGAGTTTCGCGGATTC 4023
      |||||
716 LysAlaGlyAlaHisValGluSerITrgITrPheITrgITrgITrValITrgITrLeuITr 732
      |||||
4024 GGCATCGAACCGGACATCGGCGGCAACGCGCTATTTCGTCGCAAAACGGA 4073
      |||||
```


481 ATGCCGCTTTGCACAAATTTGTGCACAGATGCACAAACCTGTTGAGATGAC 530
470 ThrValaspasnlysglyThrMetThrValThrpsProgluSerIleGI 486
531 CAGTTATATGATGGTGGAATAGCTGATTAAATTAATACCGTGATC 580
486 yIleGIhIleaspIly..... 491
581 GTGTCGAATCGAGACGACGACATATTTGGCGTCTGTGATGAAGCA 630
492AspGIhAlaIle 495
631 CCCAATACCGCGCAAGTTCATATCATATTCGAAGCCGATATTCCTGGCT 680
496 ValasnasnIuglyIuSerThrIleThrasnIyglYthrGIhIle 512
681 CGTCGGGCAATACCTTTCACAAAATGATCAGGTGGGACAGATC. 729
512 nIleasnIyasnspalThrAlasnSerGIyLysThrThrValA 529
730AACTAGGTACGGA 744
529 spGIyLyspSerThrGIyThrLysIlealaglYasnIleGIyIleVal 545
745 AAAATTAACATAGCCCATATGGTTTTCACCAACAGAGAGC..... 786
546 AsnleuaspIlySer.....LeuThrValThrGIyLysIleahIseGI 559
787TCATTTGGCGACAGTGGCTCA.....CCAATGT 814
559 yValGIuasnIleGIyaspasnIyThrValasnasnlysglyaspIleVal 576
815 TTATCTATGATGCCCAAAAGCAAAAGTGTATTAATGAGG..... 855
576 alValSeraspThrGIySerIleGIyValleuIleasnIyglIuglyAla 592
856 ..GTATTGCAACAGCAGCAACCCCTATATA..... 882
593 ThrValSerasnThrGIyaspValasnValSerasnIuAlaThrGIyph 609
882 882
609 eSerIleThrThrasnSerGIyLysValSerleuAlaglYSerMetGIuV 626
883GGAAGCAATGGCTGC 900
626 alGIyaspSerThrGIyValaspLeuasnIyasnasnSerVal 642
901 CAGCTAGTTCGTAAGATGGTTCTATGATGAATC..... 936
643 ThrleuAlaIalaspLeu.....LysValValGIyGIuLysAl 656
937TTGCTGAGATACCATTCAGTA..... 960
656 aThrGIyIleasnValSerGIyaspAlasnThrValasnIleThrGIyA 673
961TTC 963
673 snValleuValaspLysaspLysThrAlaaspSnAlaIaglYurThe 689
964 TACGAACCCATCAAAATGGAATACTTTTAAACGACATATTAATAGG 1013
690 pheaspProSerValGIyIleasnValIyGIySeraspSnasn..... 704
1014 CGCAGGAAAATGATGCCAAACATAAACACTATCTTACCTTATAGAT 1063
705ValThrleuaspGIyLysLeuThrValValSer...AspSerGIuV 719
1064 TAAAAACAGAACCGTTCATATGTTTAATGTTTCTTATCCGAGACACA 1113
719 alThrSeratGIuSerasnleuPheaspGIySerAlaglIuLysThrSer 735
1114 AGAAGACCTGTTTATATCTGACAGTGGGCTCAC..... 1149
736 GlyLeuValValIleGIyaspGIyasnThrValasnMetasnIyGIyLe 752
1150AGTTATGACCCAGACTGAATATGAGAGAAATATTTCT 1189
752 uGIuLeuIleGIyGIuLysasnAlaIleuAlaspGIySerGIuValThs 769
1190 TTATTGACAAAGAAAAGT.....GAATTGATACTTACGACAAATC 1233
769 leuLeuThrGIyLysSerYThrThrSerValIleValValSer..... 783
1234 AACCAAGCGCGCGGCTTTGATTTTGAGGGTAATTTTACGCTCCGC 1283
784GlyIuSerSerValIyLeuasnIyaspThrIleSerGI 798
1284 TAAAAACAGAACGACGACGAGC.....GGCG 1312
798 yGIuPheProleuGIyPheAlaglYValIleArgValGIuaspLysAlaL 815
1313 GCGTTCATATTCAGTATGACGACGTACCGTACTTGGAAATTAACGCGCTG 1362
815 eulGIuGIuIleGIySerGIyAlaThrleuThrMetGIu..... 827
1363 GCAACGACCGCGCTGCCAAATCGCAAGCAGCAGCTGCTGTCAA.. 1410
828 AspIleaspSerPheGIuIhIseGIyThrArgThrValGIuIleGIuAs 844
1411GCCAAAGGGGAAAACCAACGAGCTCGGACGCGG 1444
844 nleuGIyPheAlaPheValThrGIyGIuasn.....ThrThrGIyIleA 859
1445 GCGACGGTAAGTCACTTACATCAGCAGCGGACGATCAAGCAAAAA 1494
859 snSerGIyThrIleSerleuGIuIn.....AsnGIyLysasp 871
1495 CAAGCCTTAGTAATC..... 1512
872 ProAlaPProSerProIleValleuAlaThrasnIyGIySerAlaTh 888
1513GGCTTGTACGCGGACGCGGACGCTGCACACTGAATCCGATA 1555
888 rAsnAlaglYThrIleThrGIyLysValThrGIuGIuIhIseValPheA 905
1556 ATCACTTAACACC...GACAACTCTATTGGCTTTCGCGGCGACGT 1602
905 snLysTYrSerThrGIyThrSerasnSerPheIlePheasnAsnspVal 921
1603 TTGATTTGAACGGG...CATTCCTTTCGTTCCACGCGATTCAAATAC 1649
922 SerSerIleThrGIyLeuValAlaGIuInSerasnSerThrIleIleasnTh 938
1650 CGATGAACGGCGATG..... 1665
938 raspSerGIyIleIleaspLeuTYrGIyArgGIySerValGIyMetLeuA 955
1666 ..ATTGTCAACACATCAACAGACAAAGATCCACCTTACATT..... 1707
955 lalIleAlaspSerThrAlaGIuasnIyGIySerIleThrleuAspSer 971
1708ACAGCAATTAAGATTTTACTACA.. 1731
972 MetThrValaspAlaasnspThrThrAlaMetArgspIleAlaSerAs 988
1732ACCGCAATTAACAACACTTGATAGC.... 1758
988 nSerAlaIleaspPheGIyThrGIyValGIyValGIyThrAspSerTYs 1005
1759AAAAAGAAATGGCTACAC..... 1779
1005 ergIyAlaglYLysasnAlaThrAlaIleasnIuInleuGIyGIyValIle 1021
1780GTTGGTTTGGCGAGAAAGATGC 1802

```
1022 ThrIleTyrAsnAlaGlyAlaGlyMetAlaAlaTyrGlyAlaSerAsnThr 1038
      :::: :::::
1803 AACCAAAACGACGCGGCTCAATCTG.....AATTACCAAC 1840
      :::: :::::
1038 rValIleAsnGlnGlyThrIleAsnLeuGluLysAsnGlyAsnTyrAsp 1055
      :::: :::::
1841 CGAAGAAGAGGATCGCACTTA.....CTGCTTCGGGGGA 1878
      :::: :::::
1055 sPserLeuAlaIleAsnThrLeuValGlyMetAlaValTyrGluHisGly 1071
      :::: :::::
1879 ACAATTAAACGCG.....AATTACGCAAAACACG 1913
      :::: :::::
1072 ThrAlaIleAsnAspGlnThrGlyValIleAsnIleAsnValGlyThrGln 1088
      :::: :::::
1914 CAACCTGTTTTTCAGC...GGCAGACGACACCGCAGCTACATCATTT 1960
      :::: :::::
1088 yGlnAlaPheTyrAsnAspGlyThrGlyThrIleValAsnTyrGlyThrI 1105
      :::: :::::
1961 TA.....GGACGGGGTGGTCAAAATGGAA 1986
      :::: :::::
1105 lEcysThrPheGlyValGlySerGlnSerGlyAsnGluTyrAsnAsnThrAsp 1121
      :::: :::::
1987 GGTATCCCA.....CAAG 2000
      :::: :::::
1122 AspPheThrSerLeuIleTyrThrGlyGlyAspThrIleThrArgSerGln 1138
      :::: :::::
2001 AGAATCTGTGGGACACGATGATGACCGCAGCATTTAA..... 2043
      :::: :::::
1138 yGlnThrValThrLeuAsnLysSerAlaIleValThrAspLysLeuAlaG 1155
      :::: :::::
2044 .....GCGGAAACTTCATATTACGGCGGACAAAGCGGTGGTTTC 2085
      :::: :::::
1155 lYAsnValValAsnSerGlyThrLeuSerGlyAspGlnIleThrValSer 1171
      :::: :::::
2086 CGCAATGTTGCCAAAGTGAAGCGCATTTGACATTAACATCACGCCCA 2135
      :::: :::::
1172 SerGlyLeuLeuGlnAsnThrSerGlyGlyIleIleAsnAsnLeuValLy 1188
      :::: :::::
2136 ACCAGATTTCGGTGGCGACCGCATCAAGCACACAACTGTGTACAGCTT 2185
      :::: :::::
1188 sLeuAspLysGlyAlaValIleLysAsnAlaGlyValMetThrAsnAsnV 1205
      :::: :::::
2186 CGGACTGACGGGT.....CTGACAAGTTGTACGCAAAACCATTAAC 2229
      :::: :::::
1205 aLAspValSerGlyGlyIleLeuAsnAsnAlaGlyLumethr..... 1219
      :::: :::::
2230 GACGATAAAGTATGCTTCATTGACGAAGCCGACATCAGAGGCAATGT 2279
      :::: :::::
1220 .....AlaGlnIleThrMetAsnAl 1226
      :::: :::::
2280 GACGCTTGCGATCAAGCTCATTTAAATCTCACAGAGCTTGCACACATCA 2329
      :::: :::::
1226 acGly...AlaAspSerSerLeuValAsnAsnThrGly.....ThrIleA 1240
      :::: :::::
2330 AC.....GCGAAT..... 2337
      :::: :::::
1240 sNlysIleValGlnAsnAlaGlyValPheAsnAsnSerGlySerValThr 1256
      :::: :::::
2338 .....CTTAGTCGAGCGGA.....GACAGCCTACTATAC 2366
      :::: :::::
1257 GlyArgMetMetSerAlaGlyGlyValPheAsnAsnGlnThrAspGlyVal 1273
      :::: :::::
2367 GGTTCAGCGCAACGACCAACGCAACCTCAGCCCTGTGGGCAATG 2416
      :::: :::::
1273 aLleMetAlaGlyAlaIleLeuThrGly.....ThrAlaValAlaAsnA 1288
      :::: :::::
2417 CCCAAGCAACATTAAATCAAGCCACA..... 2442
      :::: :::::
1288 sNgluGlyThrTyrPasnLeuGlySerSerSerGlnGlyAsnAsnThrGly 1304
      :::: :::::
2443 .....TTAAGCGCAACACATGCGCTTGGACAAATGCTTCATTTAA 2483
      :::: :::::
```

```
1305 MetLeuGluValAsnAsnSerAlaPheAsnAsnArgGlyGluPheI 1321
      :::: :::::
2484 TCTAACCAAC.....AACCGGTACAA...AACGCGAGTCTGACGCTTT 2524
      :::: :::::
1321 eLeuAspAsnAspLysAsnAlaValHisIleAsnGlnSerGlyThrLeuT 1338
      :::: :::::
2525 CGGACAGCGCTTAAGCAACGTAAGCATTCGCACTCAACGCAATGTC 2574
      :::: :::::
1338 yAsnThrGlyHisMetAsnIleSerAsnSerSerHisAsnGlyAlaVal 1354
      :::: :::::
2575 TCCCTA..... 2580
      :::: :::::
1355 AsnMetTrpGlyGlyAsnGlyArgPheIleAsnAspGlyThrIleAspVa 1371
      :::: :::::
2581 .....GCCGATTAAGGCAG 2593
      :::: :::::
1371 lSerAlaLysSerLeuValValSerAlaAsnAsnAlaGlyAspGlnAsnA 1388
      :::: :::::
2594 TATTCCATTGTAAGAACAGCCGCTTACCGGAAAAATCAGCGCGGCAAG 2643
      :::: :::::
1388 lAspPheThrPasnGlnAsp.....AsnGlyValIleAsnPheAsnHis 1402
      :::: :::::
2644 GATACGCGATTACCTTAA..... 2664
      :::: :::::
1403 AspSerAlaSerAlaValLysValThrHisSerAsnPheIleAlaGlnAs 1419
      :::: :::::
2665 .GACAGGAAATGACGCGCGCTCGGGCAGCAATTA.....G 2701
      :::: :::::
1419 nAspGlyIleMetAsnIle...SerGlyThrGlyAlaValAlaMetGlnG 1435
      :::: :::::
2702 GCATTTTAAC.....CTTGACACGCCGCTTACGTCATTCGATTC 2742
      :::: :::::
1435 lAspLysAsnAlaGlnLeuValLysAsnGlyThrIleAsnLeuGlyThr 1451
      :::: :::::
2743 GCGTATGACACGATGCGGCA.....GGCGCCAAACCGGCAAGTCGCG 2786
      :::: :::::
1452 AlaGlyThrThrAspThrGlyMetIleGlyMetGlnLeuAspAlaAsnAl 1468
      :::: :::::
2787 AGATCGCGCGCGCGCGCTTCCGCGCTTCCTATTATCCGTTACCGCCG 2836
      :::: :::::
1468 aThrAlaAspAlaValIleGlnAsnAsnGlyThrIleAsnIlePheAlaA 1485
      :::: :::::
2837 CACTTCGGCAGATCCGTTTCAACGCGTGAAGGTAACGCAATTTG 2886
      :::: :::::
1485 sNAspSer.....PheAlaPheSerValLeuGlyThrValGlnHisVal 1499
      :::: :::::
2887 AACGCTCAGGAAACATTCGCTTATGTCGAACTTCGCTTACCGCAG 2936
      :::: :::::
1500 ValAsnAsnGlyThrValValIleAlaAspGlyValThrGly.....Se 1514
      :::: :::::
2937 CGGCAAAATTGACGTCGCGGA..... 2958
      :::: :::::
1514 rGlyLeuIleLysGlnGlyAspSerIleAsnValGlnGlyMetAsnGlyA 1531
      :::: :::::
2959 .....AGTTCGGA.....GGCACTTACACTTGGCTGTC 2988
      :::: :::::
1531 sNAsnGlyAsnSerSerGluValHisTyrGlyAspGlyThrLeuPro... 1546
      :::: :::::
2989 AACCAATACCGGCAACGACCGTAAGTCTGAGCAATTCAGCGTAGTGA 3038
      :::: :::::
1547 .....AspValProLysProAsnThrValSerValThrSe 1558
      :::: :::::
3039 AGGAAAGACACACACCGCTGTCC...GAAATCTTAATTCAACCTGC 3085
      :::: :::::
1558 rGlySerAspGluAlaGlySerMetAsnAsnLeuAsnGlyTyrValVal 1575
      :::: :::::
3086 AAACGAACAGCTGATGCGCGCATGCGGTTATCAGTTATCCGC... 3132
      :::: :::::
1575 aLglyThrAsnValAsnGlySerAlaGlyLysLeuValAsnAsnAla 1591
      :::: :::::
3133 .....AAGA 3137
      :::: :::::
1592 SerMetAsnGlyValGluIleAsnThrGlyPheThrAlaGlyThrAlaAs 1608
      :::: :::::
```


221 ArgGlnIleValArgAlaGluGlyThrAlaAsnThrValValTyr.. 236
1584 CGGGTTGGGGGGGAGGAGTTGGATTGAAAGCGCATTCGTTTCGTTCC 1633
237AlaGlyLysPglThrValHisGlyHisAlaLeu..... 248
1634 ACCGCATTAATAATACCGATGATGAGGGCGATGATTGTCAACCCATTCGA 1683
249AspThrThrLeuAsnGlyGlyTyrGlnTyrValHisAspGly 262
1684 GACAAGAAATCCACCGTTACCATTTACAGCGAATAAGATATTACTCAAC 1733
263 GlyThrAlaSerAspThrValVal.....AsnSerAsp..... 273
1734 CGGCATTAACACAACTTGATGATGCAAAAAAATGGCTTACACAGGTT 1783
274GlyTyrGlnIleValLysAsnGlyG 282
1784 GATTGGCGAGAAAGATGCAACCAACAAAGCGGCGGCTCATTCGAAAT 1833
282 LysAlaIleGlyAsnThrThrValAsnGlnLysGlyArgLeuGlnValAsp 298
1834 TACCAACCGGAAGACCGGATCGCATCTTACTGCTTCCGCGGAACAA 1883
299AlaGlyGlyThrAl 303
1884 TTAAACGGCAATATCATCGCAACAAACGCAACTGTTTTCAGCGGGA 1933
303 aThrAsnValThrLeuLysGln...GlyGlyAlaLeuValThrSerTha 319
1934 GACCGACCGCGGCGCTACATCATTTAGAAAGCGGCTGTCAAAAATG 1983
319 AlaIleThrValThrGlyIleAsnArgLeuGlyAla...PheSerValVal 334
1984 GAAGATATCCCAAGAGGAATCCGTGGGACACAGATGGATGCAC.. 2031
335 GlnGlyLysAlaAspAsnValValLeuGluAsnGlyGlyArgLeuAspVa 351
2032CGCACATTTAAAGCGAAACTTCATATTCAGGCGGAGCAACGG 2077
351 IleuThrGlyHisThrAlaThrAsnThrArgValAspAspGlyGlyThrL 368
2078 TGCTTCCCGCAAT.....GTGGCAAGTGGAAAGC 2109
368 euAspValArgAsnGlyThrAlaThrThrValSerMetGlyAsnGly 384
2110 GATTGCATTTAAGCAATCAGCCCAAGATTTCCGTTGCGACCGCA 2159
385 GlyValLeuLeuAlaAspSerGlyAlaAlaValSerGly..... 397
2160 TCAAAGCCACACAATCTGTACAGTTTCGAC..... 2190
398ThrArgSerAspGlyLysAlaPheSerIleG 408
2191TGACGGGCTGACA 2205
408 LysGlyGlyGlnAlaAspAlaLeuMetLeuGluLysGlySerSerPheThr 424
2206 AGTTGACGAAACCAATACGACGATTAAGATTCGTTTCATTCAG 2255
425 LeuAsnAlaGlyAspThrAlaThrAspThrThrValAsnGlyGlyLeuPh 441
2256 CAAGAC.....GACATCAGAGCAATGTCAGCTTGCCGATCAG 2296
441 eThrAlaAspGlyGlyThrLeuAlaGlyThrThrThrLeuAsnAsnGlyA 458
2297 CTGATTTAATCTCACAGACTGCG..... 2322
458 IleIleuThrLeuSerGlyLysThrValAsnAspThrLeuThrIle 474
2323ACACTCAACGCGCA 2336
475 ArgGlnGlyAspAlaLeuLeuGlnGlyGlySerLeuThrGlyAsnGlySe 491

2337 TCCTAGTCAGGCGGAGACACGCACTATACGTTTACGGCAACGCAACC 2386
491 rValGlnLysSerGlySerGlyThrLeuThrValSer...AsnThrThrL 507
2387 AAAAGCGCACTCAAGCTCGTGGGCAATGCCAAGCAACTTATATCA 2436
507 euThrGlnLysAlaValAsnLeuAsnGlnGlyThrLeuThrLeuAsnAsp 523
2437 GCCACATTAACGCGACATCGGCTTCGGGCAATGCTCATTTAATCT 2486
524 SerThrValThrThrAspValIleAlaGlnArgGlyThrAlaLeuSle 540
2487 AAGCAACAACGCCGTACAAACGGGCACTCG.....ACGC 2521
540 uThrGlySerThrValLeuAsnGlyAlaIleAspProThrAsnValThrL 557
2522 TTTCGACAAACGCTAAGCAACGTAAGCCATTCCGCACTACAGGCAAT 2571
557 euAlaSerGlyAlaThrTrpAsnIleProAspAsnAlaThrValGlnSer 573
2572 GTC.....TCCCTAAGCCGATAGGCAAGTATTCATTTGAAACAGCG 2615
574 ValValAspAspLeuSerHisAlaGlyGlnIleHisPheThrSerThrAt 590
2616 CTTTACCGGAAATACAGCGCGCGCAAGATACGACATTACATTAAAG 2665
590 g...ThrGlyLysPhe..... 594
2666 ACAGGATGAGACGCTGCGTGGGCAAGGATTAAGCAATTTAAACCT 2715
595ValProAlaThrLeuLysValLysAsnLeuAsnGly 606
2716 GACAACGCCACCATTTACATCATTCGCGCTATGACACGATGCGGAGG 2765
607 GlnAsnGlyThrIleSerLeuArg..... 614
2766 CGCGAAACCGCGCAGTGGGCAATGCCCGCGCGCTTCGCGCGCTT 2815
614 614
2816 CCTATTATCCGTTACGCCCACTTCGCGAGATCCGTTTCAACAGC 2865
615ValArgProAspMetAlaGlnAsnAsnAlaAspArg 626
2866 CTGACGGTAAAGCAATTTGACGTCAGGAAACATTCGCTTATATGC 2915
627 LeuValIleAspGly..... 631
2916 GGAACCTCTGCGCTACCGGACGGGCAAA.....TTGAACCTGGCGAAA 2959
632GlyArgAlaThrGlyLysThrIleLeuAsnLeuValAsnA 645
2960 GTTCGGAAGCACTTACACCTTGCGCTGTCAACATACCGCAACGAAACC 3009
645 lAsGlyAsnSerAlaSerGlyLeuAlaThrSerGlyLysGly..... 658
3010 GTAAGCTCGAGCAATTTAGCGTACTGGAAGAAAGCAACACACCGCT 3059
659IleGlnValValGluAlaIleAsnGlyAlaThrTh 670
3060 GTCCGAAATCTTAATTCACCCCTGCAAAACGAAACAGTCGATCGCGCG 3109
670 rGlnGlnGlyAla.....PheValGlnGlyAsnArgLeuGlnAlaGlyA 685
3110 CAGCGCTTATCAGCT.....ATCCGC 3132
685 lAsPheAsnTyrSerLeuAsnArgAspSerAspGlnSerTrrpLysuArg 701
3133 AAAGAGCGGAGATTCCGCTGCATATCCGCTCAAGACAAAGACAGCTTC 3182
702 SerGlnAsnAlaTyrArgAlaGluValProLeuTyrAlaSerMetLeuTh 718

```

3183 CGACAACTCGGCAAGCGGAGAAACAGAGCGCCCTTGACGGCAAAAC 3232
718 T..... 718
3233 AGGCAACTTGGCCGCAACAGCGGAGAAAAAGACACAGCGCAAGC 3282
719 .....GlnAlaMetasp 722
3283 CTGACGGCGTGGATTCGGCGCGGCGCAATGCGACCGAAAGGCAAG 3332
723 TyrAspArgIleValAlaIleGlySerArg..... 731
3333 TGTTCGCAACCGCGCGCGGAGCGAGC.....GGGAAATATCGCGCA 3376
72 .....SerHisGlnThrGlyValAsnGlyIleHisAsnSerArg 744
3377 TTATCGAGCGGAGAGAGAAAAACGGGTGCGAGCGGATTAAGACACC 3426
744 aI...ArgIleSerIleGlnGlyIleHisLeuGlyHisAspAsnGly 759
3427 GCCTTGGGAAACAGCGGAGCGGAAACCGCGCGGCTACCGCGCTT 3476
760 GlyIleAlaArgGlyAlaThrProGlnSer...SerGlySerTyrGlyPhe 775
3477 CCCCCCGCGCGCGCGCGCGGATTCGCGCAACCGCAACCGCCCAAC 3526
775 eValArgIleu..... 778
3527 CGCAACCCCAACCGCGCGCGCGCTGATCAGCGCTTATGCCAATAGCGCT 3576
779 .....GlnGlyAspLeuMet...ArgThrGluValAlaIleGly 789
3577 TTGAGT.....GAATTTCCGCAAGCGCTCAGACAGCGTTTTCG 3614
790 MetSerValThrAlaGlyValIleGlyAlaIleGlyHisSerValAs 806
3615 CGTACAGCAAGATTTGACCGCGGTTCGCGAAGACCGCGCAACGCG 3664
806 pAllyAspAspArgIleSerArgAlaGlyThrValArgAspAspAlaG 823
3665 TTTCGACAGCGCGCATCCGCGGACACCAAAACCTACCGTTGC..... 3705
823 LysLeuGlyGlyTyrLeuAsnLeuValHisThrSerSerLeuThr 839
3706 .....CAAGATTTCCGCGCGCTACCGCA 3728
840 AlaAspIleValAlaGlnGlyThrArgHisSerMetLysAlaSerSerAs 856
3729 ACAAAACGACCTGGCGCAATCGGTATGCAAAAAACCTCGGACAGCGGC 3778
856 pAsnAspAspPheArgAlaArgIleTyrPhe...LeuGlySerLeuG 872
3779 GGGTGGGACCTCGTTTTCG..... 3798
872 LuthIleuLeuProPheSerIleThrAspAsnLeuMetLeuGlnProGln 888
3799 .....CACAAACGCGGAAACCTTGCAGACGCGGATCGGCACTC 3842
889 LeuGlnTyrThrTrpGlnGlyLeuSerLeuAspArgIleLysAspAsnAl 905
3843 GGCACGCGCTTGGCGCGCGCTTTCGCGCAATACGCGCAACGCGCAGCT 3892
905 aGlyTyrValLysPheGlyHisGlySerAlaGlnHisValAlaArgAlaGlyP 922
3893 TCGACATTCGCG.....ATCAGCGCGGCGCGGCTTTTGTAGTAC 3930
922 heArgLeuGlySerHisAsnAspMetThrPheGlyGlnGlyHisThrSer 938
3931 ...GGCAGCGCTTTCAGACGCGCATCAGAGCAAAATCCGCGCGCGTGTCT 3977
939 ArgAlaProLeuArgAspSerAlaLysHisSerValSerGlnLeuProVa 955
3978 GCATTTAGCGCATTCAGGCAAGATACCGCGCGGATTTTCGCGGATTCGCA 4027

```

```

955 LAsnTrpTyrValGlnPro..... 961
4028 TCGAACCGGACATCGGCGCAACGCGCTATTGTCGCAAAAGCGGATTAC 4077
962 .....SerValIleArgThrPheSerSerArgGlyAspMet 973
4078 CGATACGAAACGCTCAATATCGCCACCGCGCGCTTCATTCACCGCTA 4127
974 Arg.....ValGlyThrSerThrAlaGly..... 981
4128 CGCGCGGCGCATTAAGCGAGATTATTCATTCAAACCGCGCAACATTT 4177
982 .....SerGlyMetThrPheSerProSerGlnAsnGlyT 993
4178 CCATACGCGCTTATTGAGCGCTGTCTATACGATCGCGCTTCGCGCAAA 4227
993 hr.....SerLeuAspLeuGlnAlaIleLeuGlnAlaArg 1004
4228 GTCCGAACGCGCGCTCAATACCGCGCTATTGCGCGAGATTTGCGCAAA 4277
1005 ValArgGlnAsnIleThrIleuGlyValAla..... 1015
4278 CCGCAGTGGCGAATGGGCGCTAAACCGCGCAATCAAGTTTCACGCTGT 4327
1016 .....GlyTyrAlaHis 1020
4328 CCGTCACGCTGCGCGCGCGCAAGGG 4353
1020 eValIleSerGlySerAlaGlnGly 1028

seq_name: SwissProt_40:OMPb_RICJA
seq_documentation_block:
ID OMPb_RICJA STANDARD; PRT; 1656 AA.
AC 006653;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Outer membrane protein B precursor (168 kDa surface-layer protein)
DE (Surface protein antigen) (Cell surface antigen 5) (Scas) (OMPb)
DE (OMPb B) (Contains: 120 kDa surface-exposed protein (Surface protein
DE antigen) (120 kDa outer membrane protein ompB); 32 kDa beta peptide).
GN OMPb.
OS Rickettsia japonica.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
OX NCBI_TaxID=35790;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TH;
RC Uchiyama T.;
RT Sequencing of the gene encoding the protein romp B of Rickettsia
RT japonica.
RT Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: THE 120 KDA SURFACE-EXPOSED PROTEIN IS A MAJOR
CC STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS A RICKETTSIAL
CC VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION (BY
CC SIMILIARITY).
CC -!- FUNCTION: THE 32 KDA BETA PEPTIDE MAY SERVE AS A MEMBRANE ANCHOR
CC (BY SIMILIARITY).
CC -!- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A S-
CC LAYER WITH HEXAGONAL SYMMETRY.
CC -!- SIMILIARITY: BELONGS TO THE RICKETTSIAE OMPA/OMPb FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC

```

DR EMBL: AB003681; BAA20138.1; ..
 DR InterPro: IPR003858; OMPA_rOMPb.
 DR Pfam: PF02708; rompa_rompb; 1.
 KW Antigen: S-layer; Cell wall.
 FT CHAIN 1 1338 120 KDA SURFACE-EXPOSED PROTEIN.
 FT CHAIN 1339 1656 32 KDA BETA PEPTIDE.
 FT DOMAIN 528 533 POLY-GLY.
 SQ SEQUENCE 1656 AA; 168097 MW; 3132A69C9DD599F CRC64;

alignment_scores:

Quality: 237.50 Length: 1561
 Ratio: 0.343 Gaps: 77
 Percent Similarity: 44.331 Percent Identity: 19.859

alignment_block:

US-09-303-518D-653 x OMPB_RICJA ..

Align seg 1/1 to: OMPB_RICJA from: 1 to: 1656

```

277 TTTTCTGTGTAATCGCGTAACGGCGTGGCGCA..... 309
    ||| ||| ||||| ||||| |||||
334 PhevalValSerValaspasnGlyAlaAlaThrIleaspGlyGlnVa 350
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
310 .TTGGCGGCGCATCAATATATTGTGACGGTGCACATAAGCGGCTATA 358
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
350 IYrAlAlAspMetValIleGlnSerAlaAsnIleasnGlyGlnVala 367
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
359 AC.....AATGTTGATTTGGTGGCGGAGGAAGCAATCCGATCAG 399
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
367 snPheArgHisIleValaspValGlyIleaspGlyThrThrIlePhe 383
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
400 CACCGCTTTCTTACCAATGTGAAAGAAATATTAATAAGCAGGAC 449
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
384 ThrAlaIleSerIleValAlaIleThrGlnAsnSerAsnPheGlyThr 400
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
450 TAACGCCCATCTTATGGCGCGGATTCATATGCCGCTTGCACAAAT 499
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
400 rAsp.....PheGly.....AsnLeuAlaAlaG 408
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
500 TTGTCACAGATGCAAGACCTGTTGAGATGACGATTATATGATGGTGG 549
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
408 IValThrValProAspThrMetThrLeuThrGlyAsnPheThrGly... 423
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
550 AAATACGCTGATTAAATAAATACCCGTATCTGTTCAATCGAGCAGG 599
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
423 ..... 423
600 CAGACAATRTTGGCGGTCTGATGAAGACGAAACCAATACCGCAAGTT 649
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
424 .....AspAlaAsnAsnProGlyAsn..... 430
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
650 CATATCATATTGCAAGCGCATATTCTTGGCTGGGCGCAATACCTTT 699
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
431 .....ThrAlaGlyValIleThrPhe 437
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
700 GCACAAAATGA.....TCAGTGTGTGACAGTCAACTTAGTAGTAG 740
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
438 AlaAlaAsnGlyThrLeuAlaSerAlaSerAlaAsnValAlaVa 454
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
741 CGAAAAA.....ATTAAACATAGCCCATATGTTTATACCAA 778
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
454 IThrAsnAsnIleThrAlaIleGlyAlaSerGlyValGlyValGln 471
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
779 CAGGAGGCTCA.....TTTGGACAGACGAGTGGCCACCA 810
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
471 euSerGlyThrHisThrAlaGluLeuArgLeuGlyAsnAlaGlySerVal 487
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
811 ATGTTATCTATGATGCCAA.....AAGCAAAATGTG 842
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
488 PheIysLeuAlaAspGlyThrValIleasnGlyValAsnGlnThrVa 504
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
843 GTTATTAATGGGTATTGCAAAACAGC..... 870
  
```

```

||||| ||||| ||||| |||||
504 ILeuValIleGlyValLeuAlaAlaGlyAlaIleThrLeuAspGlySerA 521
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
871 .....ACCCCTATATAGCAAAAGCAATGCGCTTCACCTAGTTCGT 912
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
521 IaThrIleThrGlyAspIleGlyAsnGlyGlyGlyAlaAlaLeuGln 537
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
913 AAAGATTGGTCTATGATGAA.....ATCTTGGCTGAGATAC 950
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
538 SerIleThrLeuAlaAsnAspAlaThrIysThrLeuThrLeuGlyGlyAl 554
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
951 CCATTCAGTATTTACGACACATCAATCAAAATGGAATACCTTTTAAAG 1000
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
554 AsnIleIle.....SerAlaAsnGlyGlyThrIleAsnPheG 567
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1001 ACAATATATATGCGCGCAGCAAAATCGATGCCAAACATTAACACTATTCT 1050
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
567 IAlaIleasnGlyGlyThrIleIysLeuThrSerThrGlnAsnIleVal 583
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1051 CTACCTTATAGATTAAACACGAAACCGTTCAATGTTTAAATGTTCTTT 1100
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
584 ValAspCysAspLeuAlaIleAlaThrAspGlnThrGlyValAlaSpAl 600
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1101 ATCCGAGACAGCAAGACACTGTTTATCATGCTGCAGTGGGCTC.... 1146
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
600 aserSerLeuThrAsnAlaGlnThrLeuThrIleSerGlyThrIleGly 617
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1147 .....ACAGTTATCGACCCAGACTGAATATGAGAAAT 1182
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
617 IeIleGlyAlaAsnAsnThrThrLeuGlyGlnPheAsnIleGlySerSer 633
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1183 ATTTCCTTATTTGACAAAGCAAA.....GCTGAATGATACTTAC 1223
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
634 LysThrThrLeuAsnGlyGlyAsnValAlaIleAsnIleuValIleG 650
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1224 CAGCAACATCAAC...CAAGCGCGGCGGTGTGATATT... 1260
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
650 YAsnAsnGlySerValGlnPheAlaHisAsnThrTyLeuIleThrArg 667
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1261 .....GAGGCTAATTTTACGGCTCTGCT...AAAAAC 1290
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
667 hrThrAsnAlaAlaGlyGlnGlyLysIleIlePheAsnProValAla 683
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1291 AACGAACGTGGCGCAAGCGCGGCTCATATGACATGATGCGAGTACCGT 1340
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
684 AsnAsnThrThrLeuAlaAlaGlyThrAsnLeuGlySerAlaAlaAsn 700
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1341 TACTTGGAAAGTAAC.....GGCGTGGCAACAGCAGCGCTGT 1378
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
700 oLeuAlaGlnIleAsnPheGlySerIysGlyAlaArgAlaAspThrVal 717
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1379 CCAAAATCGCGCAAGCGACCGCTG.....GTTCAAGCCAAA 1416
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
717 euAsnValIleGlyGlnGlyValAsnLeuTyAlaThrAsnIleThrThr 733
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1417 GGGGAAACCAAGGCTCG.....GTCAGCGTGGGCGAGCGTAAAGCAT 1460
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
734 AspAlaAsnValGlySerPheValPheAsnAlaGlyGlyLysAsnIleVa 750
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1461 CTTAGATCAGCAGCGGAGCATCAAGCAAAACCAAGCCTTTAGTGAAA 1510
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
750 IserGlyThrValGlyGlyGlnGlnGlyAsnLys.....PheAsnThr 765
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1511 TCGGCTGTGTCAGCGCAGGGGAGCGTGCATGATGCCGTAATACAG 1560
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
765 alAlaIleuAspAsnGlyThr...ThrValIysPheLeuGlyAsnAlaThr 780
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1561 TTCACCCCGCAAAACGCTATTTCGCTTTCGCGGCGAGCGTGGATTT 1610
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
781 PheAsnGlyAsnThrThrIleAlaAla.....AsnSerThrLeuGlnI 795
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1611 GAACGGG...CATTGCTTTCGTTCCACCGC..... 1638
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
  
```

795 eserGlyAsnTyrThrAlaAspPheIleAlaSerAlaAspGlyThrGlyI 812
1639 ..ATTCAAAATACCGATGAAGGGCGATGATGTCAACCAATCAAGAC 1686
812 leValGluPheValAsnThrGlyProIleAsnValThrIleAsnGln 828
1687AAGAAATCCACGCTTACCATTTACAGGCA 1715
829 AlaValProValAsnAlaLeuLysGlnIleThrValSerGlyProGlyAs 845
1716 TAAAGATATTACTACAACCGCAATTAACACACTTGATAGCAAAAAG 1765
845 nValValAlaAsnGlnIleGlyAsnAlaGlyAsn..... 856
1766 AAATTCCTACAACGGTTGGTTGGCCAGAAAGATGCACCAAAAGAC 1815
857TyrHisGlyAlaMetThrAsp..... 863
1816 GGGGGGCTCATCTGAATTACCAACGGAGAGACCGATCGCACTTACT 1865
864ThrIleAlaPheGluAsnSerIleuGlyAlaValLeuP 877
1866 GCTTCCGCGGAAACAAATTTAAC.....GGCAAT..... 1896
877 eleuProSerGlyIleProPheAsnAspAlaGlyAsnThrIleProLeu 894
1897 ..ATCAGCCAAACAACGGCAACTGTTTTCAGCGGCGACGACGCG 1944
894 hTleLysSerThrValGlyAsnGluThrAlaGluGlyPheSerValPro 910
1945 CACGCTTCATATTCATTAGAACGGGCGGTGTGCAAAAATGAAAGTATCCC 1994
911 SerVal.....IleValSerGlyValAspSerVal.....IleAl 922
1995 ACAAGAGAAATCGTGTGGACACGATGGATC..... 2028
922 aAspGlyGlnValIleGlyAspGlnAsnAsnIleValGlyLeuGlyLeuG 939
2029GACCGCACATTTAAAGCGGAAACTTCATATTACGCGGACAA 2073
939 lySerAspAsnGlyIleIleValAsnAlaThrThrLeuTyrAlaGlyLe 955
2074 GCGGTGTTTCCCGCATGTGGCCAAAGTGAAGGCGCATTTGACATTAG 2123
956 GlyThrIleAsnAsnAsnGlnGlyThrValThrLeuSerGlyGlyValPr 972
2124 CAATCAGCCCAAGCAGTTTCGGGTGTC.....GCAC 2155
972 onsmThrProGlyThrValTyrGlyLeuGlyThrGlyIleGlyAlaSerL 989
2156 CGCATCAAGCGCACAAATCTGTACACGTCGCGACTGCGGCGGTGACA 2205
989 yspheLysGlnValThrPheThrThrAspTyrAsnAsnLeuGlyAsnIle 1005
2206 AGTTGTACGAAAAAACCATTAACGACGATAAGATGATGCTTCA.... 2250
1006 IleAlaThrAsnThrThrIleAsnAspGlyValThrValThrThrGlyG 1022
2251TTGAGCAAGACGACATCAGACGATGAGCAGATGACACC 2284
1022 yIleAlaAlaGlyGlyIleAlaGlyThrAspPheAspGlyLysIleThr 1038
2285 TTGCGGATCAGCGCTCATTTAAATTCACAGGACTTGCCACACTCAAGGC 2334
1039LeuGlySerValAsnGly 1044
2335 AATCTTAGTGA.....GGCGAGACAGCCACTA 2363
1045 AsnAlaAsnValAlaPheAlaAspGlyIlePheSerAsnSerThrSerMe 1061
2364 TACGGTTACGCGCAACGCCAAACGCAACTCGACGCCGCGGGCA 2413
1061 tIleValThrThrLysAla..AsnAsnGlyThrValThrValThrLeuGly 1077
2414 ATGCC..... 2418
1077 snAlaPheValGlyAsnIleGlyAspSerAspThrProValAlaSerVal 1093
2419CAAGCAACATTTAATCAAGCCACATTTAAACGGCAAC.....AC 2456
1094 ArgPheThrGlySerAsnAsnGlyAlaGlyLeuLysGlyAsnIleTyrSe 1110
2457 ATCGGCTTGCGCATGCTTCATTATCTA.....ACCAACAGC 2497
1110 rGlnValIleAspPheGlyThrTyrAsnLeuGlyIleValAsnSerAsnV 1127
2498 CCGTACAAACGCGACGCTGACGCTTCCGACACGCTAAGCAAAACGTA 2547
1127 alIleLeuGlySerThrThr..... 1134
2548 AGCCATTCCGCACTACAGCGCAATGTCTCCTAGCCGATAGCGATATT 2597
1135AlaIleAsnGlyLysIleAspLeuLeuThrAsn..... 1145
2598 CCATTTTGAACAGCGCGCTTACCGGAAATACAGCGGCGGCAAGATA 2647
1146ThrIleuThrPheAlaGlyGlyThrSerT 1155
2648 CGGCTTACACTTAAAGACAGAGAAATGACGCGCTGCGGCGACAGAA 2697
1155 hTrrpGlyAsnAsnThrSerIleGluThrThrLeu.....Thr 1167
2698 TTAGGCAATTTAACTTCGACAAAGCGACCATTTACGCTCAATTCGCCCTA 2747
1168 LeuAlaAsnGlyAsnIleGlyHisIleValIle..... 1178
2748 TCGACAGATGCGCGAGCGCGCAACCGCACTGCGGAGATGCGCGC 2797
1179AlaGluGlyAlaGlnValAsnAla..... 1186
2798 GCCGCGTTCGCGCGCTCCCTCAATTATCCGTTACGCCGCAACTTCGGCA 2847
1187ThrThrThrGlyThrThrThrIleAsnValGlnAspAsnAla 1200
2848 GAATCCGTTTCAAC.....ACGCTGACGCTA..... 2874
1201 AsnAlaAsnPheSerGlyThrGlnThrTyrThrLeuIleGlnGlyVal 1217
2875AACGGCAATTTGAACGT.....CAAGCAACAT 2902
1217 aArgPheAsnGlyThrLeuGlyGlyProAsnPheThrValThrGlySerA 1234
2903 TCCGCTTTATGTGGAACCTTCCGCTACCGCGCAACATTTGAAGCTG 2952
1234 snArgPheValAsn.....TyrGlyLeuIleAlaGala 1244
2953 GCG.....GAAAGTTCGAAGGCACTTA 2975
1245 AlaAsnGlnAspTyrValIleThrArgThrAsnAlaGlnAsnIleVal 1261
2976 CACCTTGCGTGCACATATC.....GGCAAGAACCCGTAAGTCTCG 3019
1261 lThrAsnAspIleThrAsnSerProPheGlyGlyAlaProGlyValGlyG 1278
3020 AGCATTTGACGGTATGGAAGAAAGACACACACCGCTGTCCGAAAT 3069
1278 lAsnValThrThrPheValAlaAsnAlaThrAsnThrAlaLaryAsnAsn 1294
3070 CTTAATTTCAACCCGCAAAACGAACACGTCGATGCGGCGCATGGCTTA 3119
1295 Leu..... 1295
3120 TCACCTTATCCGCAAGACGCGAGTTCGCTGCAATATCGGCTCAAG 3169
1296LeuLeuAlaLysAsnSer..... 1301

```

3170 AACAGAGCTTCCGACAACTCGCAGCGGGA.....GAA 3207
      |||
1302 .....AlaaspserrAlaasnPhValGlyThrIleValThrAsp 1314
      |||
3208 ACAGAGCGCGCTTGACGGCAAAAGCGGCAACACTGCCGCAACACAA 3257
      |||
1315 ThrSerAlaIleThrAsnAlaGlnLeuAspValAlaIleAspIleG 1331
      |||
3258 GCGGCAAAAGACACGCGCAAGCCTTGACGGCTGATTCGGCGCGGC 3307
      |||
1331 nAlaGlnLeuGlyAsn.....ArgLeuGlyAlaLeuArgIleGlyT 1346
      |||
3308 GCATATCCACCCGAAAGCGAAGAGTTCGCGCAACGCGCGCGAGCA 3357
      |||
1346 hProGlnMetValGlySerGlnAlaGlyAlaIleProAlaIleVal 1362
      |||
3358 GCGCGGCAAAATGCC.....GGCATATGACAGCGCA 3389
      |||
1363 AlAGlyAspGlnAlaValAspAsnValAlaIleGlyIleTrp..... 1376
      |||
3390 GGAAGAAAAAACGGGTGCGAGCGGATTAAGACACCGCCTTGCGA 3439
      |||
1376 ..... 1376
3440 AGCGCGAAGCGAAACCGCGCGCTACCAACCGCTCCCGCGCGCGC 3489
      |||
1377 .....AlaIysProPheIleThrAspAlaHis 1385
      |||
3490 CCGCGCGCGCGGATTTGCCGACCGCGCGCAACCGCAACCGCAAC 3539
      |||
1386 GlnSerLysLys..... 1389
      |||
3540 GCAGCGCGACTGATCAGCGCTTATGCAATAGCGTTTGAATATTT 3589
      |||
1390 .....GlyGlyLeuAlaGlyTyrL 1396
      |||
3590 CCGCGACGCTCAACAGCGTTTGGCGGTACAGAGCAATTCGACCGC 3639
      |||
1396 yAlaIleThrThrGlyIleValIleGlyLeuAspThrLeu..... 1409
      |||
3640 TTGGCGAAGACCGCGCGCGCGTTCGACAGCGCGATC.....CG 3683
      |||
1410 .....AlaAsnAsnAlaMetIleGlyAlaAlaIleGlyIleThrLys 1425
      |||
3684 GGACACCAACAC.....TACGTTCCCAAGAT..... 3711
      |||
1425 rAspIleLysHisGlnAspTyrLysGlyAspLysThrAspValAsn 1442
      |||
3712 .....TTCGCGCGCTACCGCGCAACCGCGCGCAATTCGCT 3753
      |||
1442 LysPheSerPheSerLeuTyrGlyAlaGlnPheValGlnAsnPhPhe 1458
      |||
3754 ATGACAGAAAAACCTCGCGAGCGCGCGCTCGGATCCTGTTTGGCA 3803
      |||
1459 AlaGln.....GlySerAla.....IlePheSerLeuAs 1468
      |||
3804 CCGGACCGGAAACACC.....TTCGACGACGCGCATTCGCGCAAC 3840
      |||
1468 nGlnValLysAsnLysSerGlnArgTyrPhePheAspAlaAsnGln 1485
      |||
3841 ..TTCGCGACGCGCTTCGCGCGAT.....GCCGTTTTCGGCGCA 3876
      |||
1485 etSerLysGlnIleAlaIleGlyAsnTyrAspAsnMetThrPheGly 1500
      |||
3877 TACGGCATCGCGAGTTCGACATCGGC..... 3903
      |||
1501 .....GlyAsnLeuThrValGlyTyrAspTyrAsnAlaMetGlnI 1514
      |||
3904 .....ATCAGCGCGCGCGCGGTTTATGACGCGCGAGCTTCAGAC 3946
      |||
1514 yAlaLeuValThrProMetAlaGlyLeuSerTyrLeuLysSerSerAs 1531
      |||
3947 GCATGAGAGGCAAAATCCGCGCGCGCTGCTGCAATACGCAATTCAG 3996
      |||

```

```

1531 LuAsnTyrLysGlnThrThrValAlaAsnLysGlnValAsnSer 1547
      |||
3997 AGATACCGC.....GCAGGT.. 4011
      |||
1548 LysPheSerAspArgThrAspLeuIleValGlyAlaLysValAlaGly 1564
      |||
4012 .....TTCGCGGATTCGGCGATTCGAGCGCGCGCGCGCGCGC 4051
      |||
1564 YThrMetAsnIleThrAspPheAlaValTyrProGlnValHisAla... 1579
      |||
4052 GCTATTTGCTCCAAAAGCGGATTACGATACGAAACGTCATATATGCC 4101
      |||
1580 ..PheValValHisLysValThrGlyAlaGlyLeuSerLysThrGlnSer 1595
      |||
4102 ACCCGCGC.....CTTGCAAT 4118
      |||
1596 LeuAspGlyGlnValThrProCysIleSerGlnProAspArgThrAla 1612
      |||
4119 CAACCGCTACCGCGCGCGCGCATTAAGCGAGATTA.....TCATTCAA 4162
      |||
1612 sThrSerTyrAsnLeuGlyLeuSerAlaSerIleArgSerAspAlaLys 1629
      |||
4163 CCGCGCAACACATTTCCATCAGCGCTTATTTGACCTGCTATACCGAT 4212
      |||
1629 etGluTyrGlyIleGlyTyrAspAlaGlnIleAlaSerLysTyrThr... 1644
      |||
4213 CCGCGCTCCGCGCAAGTCGCGAGCGCGTCAT 4245
      |||
1645 AlaHisGlnGlyThrLeuLysValArgValAsn 1655
      |||

```

seq_name: SwissProt_40:OMP_RICPR

seq_documentation_block:

```

ID OMP_RICPR STANDARD; PRT; 1643 AA.
AC Q53020; Q9ZCM0;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Outer membrane protein B precursor (168 kDa surface-layer protein)
DE (surface protein antigen) (cell surface antigen 5) (Scas5) (OMPb)
DE (OMP B) [contains: 120 kDa surface-exposed protein (Surface protein
antigen) (120 kDa outer membrane protein ompB); 32 kDa beta peptide].
GN OMPB OR SPAP OR SPA OR RP704.
OS Rickettsia prowazekii.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
OX NCBI_TaxID=782;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN=BREINL;
RX MEDLINE=91045972; PubMed=2122457;
RA Carl M., Dobson M.E., Ching W.M., Dasch G.A.;
RT "Characterization of the gene encoding the protective paracrystalline-
RT surface-layer protein of Rickettsia prowazekii: presence of a
RT truncated identical homolog in Rickettsia typhi."
RL Proc. Natl. Acad. Sci. U.S.A. 87:8237-8241(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=BREINL;
RA Moron C.G., Yu X.J., Walker D.H.;
RT "Sequence analysis of ompb of Rickettsia prowazekii."
RL Submitted (JUN-1999) to the EMBL/Genbank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=MADRID E;
RX MEDLINE=99039499; PubMed=9823893;
RA Andersson S.G.E., Zomorodipour A., Andersson J.O.,
RA Sichteritz-Ponten T., Alsmark U.C.M., Podowski R.M., Naeslund A.K.,
RA Eriksson A.-S., Winkler H.H., Kurland C.G.;
RT "The genome sequence of Rickettsia prowazekii and the origin of
RT mitochondria."
RL Nature 396:133-140(1998).

```

```
RN [4]
RP PARTIAL SEQUENCE.
RC STRAIN=BREINL;
RX MEDLINE=92114896; PubMed=1370573;
RA Ching W.M., Carl M., Dasch G.A.;
RT Mapping of monoclonal antibody binding sites on CNBR fragments of
RT the S-layer protein antigens of Rickettsia typhi and Rickettsia
RT prowazekii.
RL Mol. Immunol. 29:95-105(1992).
RN [5]
RP IDENTIFICATION OF CLEAVAGE SITE.
RX MEDLINE=92104668; PubMed=1729180;
RA Hackstadt T., Messer R., Cieplak W. Jr., Peacock M.G.;
RT Evidence for proteolytic cleavage of the 120-kilodalton outer
RT membrane protein of rickettsiae: identification of an avirulent mutant
RT deficient in processing.
RL Infect. Immun. 60:159-165(1992).
CC -1- FUNCTION: THE 120 KDA SURFACE-EXPOSED PROTEIN IS A MAJOR
CC STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS A RICKETTSIAL
CC VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION.
CC -1- FUNCTION: THE 32 KDA BETA PEPTIDE MAY SERVE AS A MEMBRANE ANCHOR.
CC -1- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A S-
CC LAYER WITH HEXAGONAL SYMMETRY.
CC -1- SIMILARITY: BELONGS TO THE RICKETTSIAE OMPA/OMP FAMILY.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
```

```
283 hrcGluPheAsnValThrGlyThrLeuGlyGlyAsn..... 294
805 TCACCAATGTTATCTATATGATGCCCAAGCAAAAGTGSTTAATAATGG 854
295 .....:|||||:|||||:|||||:|||||
855 GGTATGCAA.....LeuYsG1 297
297 yllelleGluLeuAsnThrAlaValAlaGlyLysLeuIleSerGln 314
865 .....ACAGGCAACCCCTATATGCAAAAAGCAATGCTCCAGCTAGT 909
314 lylGlyAlaAlaAsnAlaValIleGlyThrAspAsnGlyAlaGlyrrgla 330
910 CGTAAGATTTGTTCTATGATGAATCTTGGTGGAGATACCATTCAGT 959
331 .....AlaGlyPheIleValSerVa 337
960 ATTCTACGAACACATCAAAATGGAAATACCTTTTAAACCAATATA 1009
337 .....:|||||:|||||:|||||:|||||
1010 ATGGCGCA.....GCAAAATGATGCCCAAAACATAAACACTTCT 1050
341 snAlaIatThrIleSerGlyGlnValTyrAlaLysAsn..... 353
1051 CTACCTTATGATTAATAAACACAGACCGTTCAATGTTATGTTCTTT 1100
354 .....:|||||:|||||:|||||:|||||
1101 ATCCGAGACAGCAAGAACCTGTTATATGCTGAGTGCGGGGTACACA 1150
363 yGlnValThrPheGluHisIleValAspValGlyLeuGlyThrThra 380
1151 GTTATGACACCCACATGTAATGAGAAATTTTCCTTATATGACANA 1200
380 snPheLys.....Thr 383
1201 GGAAGAGTGAATGTACTTACCAACACATCAACCAAGCGCG..... 1245
384 AlaSperLysValIleIleThrGluAsnSerAsnPheGlySerThras 400
1246 .....GGCGGTTTG.....TATTGG 1261
400 nPheGlyAsnLeuAspThrGlnIleValValProAspThrLysIleLeuL 417
1262 AGGCTATTCTTACGCTGCTGCTAAACAAACGAA..... 1296
417 yslGlyAsnPheIleGlyAspValLysAsnGlyAsnThrAlaGlyVal 433
1297 ...ACGTGGCAAGCGCGCGCTTCATATCACT.....GATGGCAG 1334
434 IleThrPheAsnAlaAsnGlyAlaLeuValSerAlaSerThrAspProAs 450
1335 TACCGTTACTTGGAAAGTAACGGC..... 1359
450 nIleAlaValThrAsnIleAsnAlaIleGluAlaGlyValAlaGlyVal 467
1360 .....GTGGCAACGACCGCTGTCC..... 1380
467 alGluLeuSerGlyIleHisIleIleAlaIleuAsnGlyLeuAsnGlyGly 483
1381 .....AAATCGCAAGAGCAGCGCTGCTGTTCAACCCAAAGGGGA 1421
484 SerIlePheLysLeuAlaAspGlyThrValIleIleAsnGlyProValAsnG1 500
1422 AAC.....CAAGCTGCGTACGCTGG 1444
500 nAsnAlaLeuMetAsnAsnAlaLeuAlaIaGlySerIleGlnLeu 516
1445 GCGACGCTAAAGTCATCTTAGATCAGCAGCGGACGACGCAAGCAAAAA 1494
517 ..AspGlySerAlaIleIleThrGlyAspIleGlyAsnGlyGlyValAsn 532
```


1495 CAAGCCTTTAGTGAATGGCGTGGTGCAGCGGCGAGGGGACGGG...CA 1541
1533 A1a1a1eUglnh1s1eThrLeuAlaAsnAspAlaSerLys1leUal 549
1542 ACTGAATGGCGATTAATCAGTTCAACCCGAC.....AAACTGATT 1582
549 aleuAspGlyAlaAsn1le1eGlyAlaAsnValGlyAla1leHisP 566
1583 TCGGCTTTCGCGGACGTTTGGATTGAAAGGGGACATTCGTTTCGTT 1632
566 heGlnAlaAsnGlyThr1le..... 573
1633 CACCGCATTCAAATATCCGATGAGAGGGCGATGATTGTCAACCAATCA 1682
574 ..LysLeuThrasnThrlasnsn1leVal.....AsnH 586
1683 AGACAAGAATCCACCGTTACCATTAACGCAATTAAGATTACTATCA 1732
586 eAspLeuAsp1leThrThraspLysThrlGlyValAlaAspAlaSer1e 603
1733 CCGGCATTAACAACACTTGGATTGACAAAAAAGAAATGCTACACAGT 1782
603 eutHrasnsnGlnThrLeu.....ThrlAsnGly 613
1783 TGGTTTGGCGAGAAAGATGCA...ACCAAAACGAGCGGCGGCTCATCT 1829
614 Ser1eGlyThrValValAlaAsnThrLysThrlAsnGlnLeuAsn1l 630
1830 GAATTATCCAAACCGAAGACGATCGACTTACTGTCTTCCGCGCGAA 1879
630 eGly.....SerSerLysThrlleUeAsnAlaGlyAspV 642
1880 CAATTTAAACGGCATATACGCAACAAACGGCAACTGTTTTCAGC 1929
642 a1a1a1eAsnGlnLeuVal1leGlnAsnAsnGlySerValGlnLeuAsn 658
1930 GCGAGA.....CCGACACCGCAGCGCTACAT..... 1956
659 HisAsnThrLysLeu1leThrLysThrlAsnAlaAlaAsnGlnGly 675
1956 1956
675 n1le1leValAla1a1aAspProLeuAsnThrasnThrThrlleUa1aspG 692
1957CATTTAGAGACGGGCTGCTCAAAAATGAA..... 1986
692 LysThrasnLeuGlySer1aGlnAsnProLeuSerThrlleHisPheAla 708
1987GGTATCCCAAGAGCAAT 2006
709 ThrlLysAla1a1aAsnAlaAspSer1leUeAsnValGlyLysGlyValAs 725
2007 CGTGTGGGACAACGATTGGATGACCGACCATTTAAACGGGAAACTTC 2056
725 n1eUtyrAlaAsnAsn1leThrThrasnAspAlaAsnValGlySerLeuH 742
2057 ATATTGAG...GGCGACAAGCGGTGTTTCCGCAATGTTGCCAAAGT 2103
742 1sPheArGserGlyGlyThrSer1leValSerGlyThrValGlyGly6n 758
2104 GAAGGCGATTGGCATTTAGCATATCAACGCCCAAGCAGTTTCGGTTCGC 2153
759 GlnGly.....HisLysLeuAsnAsnLeu1leUeAspAsnGly..... 771
2154 ACCGCATCAAAAGCCACACATCTGTACACGTTCCGACTGGAGGGGTCTGA 2203
772ThrThrValLysPheLeuGlyAspT 780
2204 CAAGTTGTATCCGAAAAACCATTCACGACGATTAAGTATGCTTCATG 2253
780 hrThrPheAsnGlyGlyThrLys1leGlyGlyLysSer1leUeGln1le 796

```

1061 ThrValThrTyrLeuGlyAsnAlaLeuValSerAsnIleGlySerLeuAs 1077
1077 pnhProValAlaSerValArgPheThr.....GlyAsnAspSer. 1090
3053 CACCGCTGCGAATACTTAATTCACCCCTCAAAAGCAACACGCTGCAT 3102
1091 .....GlyAlaGlyLeuGlnGlyAsnIleTyrSerGlnAsnIleAsp 1104
3103 GCGCGCGCATGGCGCTTATCAGCTTATCCGCAAGACGGCGAGTCCGCT 3152
1105 PheGlyThrTyrAsnLeuThrIle..... 1112
3153 GCATATCCGCTCAAAAGAACAGAGCTTTCGCAAAAGTCCGCGCAGGGCG 3202
1113 ..... 1121
3203 GAGAAACAGAGAGCGCCCTTGACGCAAAAGCAACACTGGCCGCCAAA 3252
1121 LysIYThr.....ThrAlaIleAsnGlyGluIle..... 1130
3253 CAAACAGCGGAAAAAGACACACCGCAAGCCCTTGACGCGCTGATTGCGGC 3302
1131 .....AspLeuLeuThrAsnAsnLeuIlePheAlaAs 1141
3303 CCGGCGCATGGCCACCGAAGAACAGAGTGTTCGCAACCGCGCGCGCC 3352
1141 nGlyThrSerThrTppGlyAspAsnThrSerIleSerThrThrLeuAsnV 1158
3353 AGCAGCGCGGGAATAATCCGCGCATTTGACCGCGAGAGAGAAAAA 3402
1158 alSerSerGlyAsnIleGlyGlnValIleAlaGlnAspAlaGlnVal 1174
3403 CCGGTCAGCGGGAATAAGACACCGCTTGCGAAGACCGCGAAGCGGA 3452
1175 AsnAlaThrThrThrGlyThrThrThrIleLysIleGlnAspAsnAlaAs 1191
3453 AACCGCG.....CGGCTACACCGCGCTTCCCGCGCGCGCGCGCGCC 3499
1191 nAlaAsnPheSerGlyThrGlnAlaThrThrLeuIleGlnGlyAla 1208
3500 GGGATTTCGCCCAACCGCAGCCCAACCGCAACCGCAGCGCGAC 3549
1208 Tg..... 1208
3550 CTGATCAGCCGTTATGCCAATAGCGGTTTGAGTATTTCCGCGAGCT 3599
1209 .....PheAsnGlyThrLeuGlyAlaProAsnPheAlaValThrG 1222
3600 CAACAGCGTTTTCGCGCGTACAGAGAGATTTGGACCGCGTGTTCGCGAG 3649
1222 YSerAsnIlePhe.....ValLysTyrGluLeu..... 1231
3650 ACCGCGCGAAGCGCTTTGACAAGCGCATCCGCGACACCAAACTAC 3699
1232 .....IleArgAspSer..... 1235
3700 CGTTCGCAAGATTTCGCGCGCTACCGCAACAAACGACCTGCGCAAT 3749
1236 .....AsnGlnAspTyr..... 1239
3750 CGGTATGCGAATAAACCCTCGCAGCGGCGCGTGCATCTGTTTTCG 3799
1240 .....ValLeuThrArgT 1244
3800 ACAACCGGACCGGAACACCTTGACAGCGCATCGCAACTGCGCAG 3849
1244 hrAsnAspValLeuAsnValValThrThrAlaValGlyAsnSerAlaIle 1260
3850 CTTCGCCACAGCGTGGCTTTCGGGCATTCAGCGATCGCA.....GGTTCGA 3896

```

```

1261 AlaAsnAlaProGlyVal..SerGlnAsnIleSerArgCysLeuGlnSerT 1277
3897 CATCGGCATCAGCGCGCGCGGCTTTTACTACG..... 3931
1277 hrAsnThrAlaAlaTyrAsnAsnMetLeuLeuAlaLysAspProSerAsp 1293
3932 .....GCAGCCCTTTCAGACGCAATCAGAGCAAGCAAAATCCGCGCGCGCTG 3978
1294 ValAlaThrPheValGlyAlaIleAlaThrAspThrSerAla..... 1308
3979 CATTCAGCGCATTCAGCAGATACCGCGAGTTTCGGCGGATTCGCAT 4028
1309 ..ValThrThrValAsnLeuAsnAspThrGlnLysThrGlnAspLeuAs 1325
4029 CGAACCGCATATCGCGCAACGCGCTATTTCGTCCAAAAGCGGATTCAC 4078
1325 erAsnArg..LeuGlyThrLeuArgTyrLeuSerAsnAla..... 1337
4079 CATACGAAACGTCATATTCGCCACCGCGCGCTTCATTCACACGCTAC 4128
1338 .....GluThrSerAspValAla.....GlySerAlaThrGlyAlaVal 1350
4129 CCGCGCGGCATTTAAGCAGATTTCATTCAAAACGCGCGCACACATTTTC 4178
1351 SerSerGlyAspGlnAlaGlnValSerTyr.....GlyValThr 1363
4179 CATACGCGCTTATTGAGCGCTGCTATACCGATTCGCGCTCCGCG.... 4224
1363 pAlaLysProPheTyrAsnIleAlaGlnGlnAspLysGlyGlyIleA 1380
4225 .....AAAGTCGACGCGCG..... 4239
1380 laGlyTyrLysAlaLysThrThrGlyValValGlyLeuAspThrLeu 1396
4240 .....GTCAATACGCC 4250
1397 AlaSerAspAsnLeuMetIleGlyAlaAlaIleGlyIleThrLysThrAs 1413
4251 CGTATTGGCGCAGATTTCGGCAAAACCGCGAGTGGGGAATGGCGGTAA 4300
1413 pIleLysHisGlnAspTyrLysLys.....GlyAspL 1424
4301 ACGCCGAATCAAGATTTCACGCTGCCCTCCACGCTGCCCGCGCAAG 4350
1424 YThrAspIleAsnGlyLeuSerPheSerLeuTyrGlySerGlnGlnLeu 1440
4351 GGGCGCATTTGAGACGCGACGACGCGCG 4380
1441 ValLysAsnPheAlaGlnGlyAsnAla 1450

seq_name: SwissProt_40:OMPb_RICCN

seq_documentation_block:
ID OMPb_RICCN STANDARD; PRT; 1655 AA.
AC Q9KKA3; Q9KKA3; Q9KKA3; Q9KKA3;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last sequence update)
DE Outer membrane protein B precursor (168 kDa surface-layer protein)
DE (Surface protein antigen) (Cell surface antigen 5) (Scab) (OMPb)
DE (omp B) (contalins: 120 kDa surface-exposed protein (Surface protein
DE antigen) (120 kDa outer membrane protein ompB); 32 kDa beta peptide).
GN OMPb OR R01085.
OS Rickettsia conorii.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
OX NCBI_TaxID=781;
RX MEDLINE=21442074; PubMed=11557893;
RX Ogata H., Audic S., Renesto-Audiffren P., Fournier P.-E., Barbe V.,
RX Samson D., Roux V., Cossart P., Weissenbach J., Claverie J.-M.,

```

RA Roullet D.;
 RT "Mechanisms of evolution in Rickettsia conorii and R. prowazekii";
 RL Science 293:2093-2098(2001).
 RN [2].
 RP SEQUENCE OF 33-1649 FROM N.A.
 RC STRAIN-Indian tick typhus, and Malish 7;
 RX MEDLINE-20393643; PubMed-10939649;
 RA Roux V., Roullet D.;
 RT "Phylogenetic analysis of members of the genus Rickettsia using the
 RL gene coding the outer-membrane protein ompB (ompB)."
 RN Int. J. Syst. Evol. Microbiol. 50:1449-1455(2000).
 RP SEQUENCE OF 353-1655 FROM N.A.
 RC STRAIN-Malish 7;
 RA Stenos J., Walker D.;
 RT "The rickettsial outer membrane protein A and B genes of Rickettsia
 RL australis, the most divergent Rickettsia of the spotted fever group";
 Submitted (May-1999) to the EMBL/Genbank/DBJ databases.
 CC -1- FUNCTION: THE 120 KDA SURFACE-EXPOSED PROTEIN IS A MAJOR
 CC STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS A RICKETTSIAL
 CC VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION (BY
 CC SIMILARITY).
 CC -1- FUNCTION: THE 32 KDA BETA PEPTIDE MAY SERVE AS A MEMBRANE ANCHOR
 CC (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A S-
 CC LAYER WITH HEXAGONAL SYMMETRY (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE RICKETTSIA OMPA/OMP FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: AE008659; AAL03623.1; -
 DR EMBL: AF123721; AAF34124.1; -
 DR EMBL: AF123726; AAF34129.1; -
 DR EMBL: AF149110; AAD39533.1; -
 DR InterPro: IPR003858; OMPA_rOmpB.
 DR Pfam: PF02708; rOmpA_rOmpB; 1.
 KW Antigen; S-layer; Cell wall; Complete proteome.
 FT CHAIN 1 1334 120 KDA SURFACE-EXPOSED PROTEIN.
 FT VARIANTS 1335 1655 32 KDA BETA PEPTIDE.
 FT VARIANTS 61 61 P -> A (IN STRAIN INDIAN TICK TYPHUS).
 FT VARIANTS 75 75 G -> S (IN STRAIN INDIAN TICK TYPHUS).
 FT VARIANTS 78 78 K -> N (IN STRAIN INDIAN TICK TYPHUS).
 FT VARIANTS 251 251 V -> A (IN STRAIN INDIAN TICK TYPHUS).
 FT VARIANTS 413 413 N -> D (IN STRAIN INDIAN TICK TYPHUS).
 FT VARIANTS 959 959 I -> V (IN STRAIN INDIAN TICK TYPHUS).
 FT VARIANTS 988 988 A -> T (IN STRAIN INDIAN TICK TYPHUS).
 FT VARIANTS 1139 1139 R -> L (IN STRAIN INDIAN TICK TYPHUS).
 FT VARIANTS 1339 1339 KD -> GH (IN REF. 3).
 FT CONFLICT 353 354 F -> S (IN REF. 3).
 FT CONFLICT 776 776 E -> D (IN REF. 3).
 FT CONFLICT 1159 1159 G -> S (IN REF. 3).
 FT CONFLICT 1177 1177 H -> R (IN REF. 3).
 FT CONFLICT 1492 1492
 SQ SEQUENCE 1655 AA: 168342 MW: E49E19377D5FCE37 CRC64;

alignment_scores:
 Quality: 237.00 Length: 1347
 Ratio: 0.337 Gaps: 65
 Percent Similarity: 45.443 Percent Identity: 18.293

alignment_block:
 US-09-303-518d-653 x OMPB_RICCN
 Align seg 1/1 to: OMPB_RICCN from: 1 to: 1655

679 CTCGTCGGTGCATACCTTTGTCACAAATGATGAGTGGTGGCAGAGT 728
 :::::::::: :::::
 :::::::::: :::::

1199 IleuenglyaspasnaAlaValIleValasnGlyAla...AsnGlyThrLe 214
 729 CAACCTTAGGTAGCGAAAAATTAAACATAGCCCATGATGGTTTTCACAA 778
 ::::: ::::: ::::: :::::
 214 uasnValThrasnGlyPheIleGlyValSerSerLysSerPheAlaThrV 231
 779 CAGACAGC...TCATTTGGCGACAGTGGCTCACCAATGTTTATCTATGAT 825
 ::::: ::::: ::::: :::::
 231 AlaasnValIleasnIleGlyAspGlyGlnGlyIleMetPheasnThrAsp 247
 826 GCCCAAAAGCAAAAGTGGTTA..... 846
 ::::: ::::: ::::: :::::
 248 AlaaspasnValasnThrLeuasnLeuGlnAlaasnGlyAlaThrIleTh 264
 847 ATTAATGGGGTATTGCAACAGCAACCCCTATATAGAAAGCAATG 895
 ::::: ::::: ::::: :::::
 264 rPheasnGlyThrAspGlyThrGlyValLeuLeuSerLysAsnA 281
 896 GCTTCAGCAGTAGTTCGAAAGATTGCTTCATGATGCAAACTTTGCTGGA 945
 ::::: ::::: ::::: :::::
 281 Ia.....AlaAlaThrAspPheasnValThrGlySerLeuGlyGly 294
 946 GATPACCATTCAGTA.....TTCAGCAACCAACATCAAAATGGAAATA 989
 ::::: ::::: ::::: :::::
 295 AsnLeuGlyGlyIleIleGlyPheasnThrValAlaValasnGlyGln 311
 990 CTTTTTTAC.....GACATATATATGCGC 1015
 ::::: ::::: ::::: :::::
 311 uLysAlaasnAlaGlyAlaasnAlaValIleGlyThrAsnasnGlyA 328
 1016 CAGGAAAAATCGATGGCAACATTAACATATTCCTACCTTATGANTTA 1065
 ::::: ::::: ::::: :::::
 328 IaGlyArg.....AlaAlaGlyPheVal 1336
 1066 AAAACACAGAACCGTCAATGTTTATGTTCTTATCGAGACAGCAAG 1115
 ::::: ::::: ::::: :::::
 337 SerValaspasnGlyValAlaThrIleaspGlyGlnValTyrAlaL 353
 1116 AGAACCTGTTTATCATGCTGCAGAGGGGCTCAC.....ACTATTCGAC 1159
 ::::: ::::: ::::: :::::
 353 SAspMetValIleGlnSerAlaasnAlaValGlyGlnValasnPheArg 370
 1160 CCAGACTGGAATATATGA...GAAATATTTCTTATGACAAAGGAA 1206
 ::::: ::::: ::::: :::::
 370 IsIleValaspValGlyThrAspGlyThrAlaPheLysThrAlaA 386
 1207 GGTGAATGATTACTTACAGCAACATCAACCAAGCGCG..... 1245
 ::::: ::::: ::::: :::::
 387 SerLysValAlaIleThrGlnasnSerasnPheGlyThrThrAspPheG 403
 1246GGCGGTTTGATTTTGAGGGTA 1267
 403 yasnLeuAlaIaGlnIleIleValPheasnThrMetThrLeuasnGlyA 420
 1268 ATTTACGGCTCTGCGCTTAACAAAGCAAGAGCGGCGGGCGGT 1317
 ::::: ::::: ::::: :::::
 420 snPheThrGlyaspAlaSerasnProGlyasnThrIaGly..... 433
 1318 CATATCAGTGAATGCGAGTACGTTACTTGAAGACTAAACGCG..... 1359
 ::::: ::::: ::::: :::::
 434ValIleThrPheaspAlaasnGlyThrLeuAl 444
 1360GTGGCAACAGCAGCGCTGCGCA 1381
 444 aserAlaSerAlaaspAlaasnValAlaValThrAsnAsnIleThrAla 461
 1382 AAATGCGCAAAAGCAGCGCTGCTGATCAAGCCAAAGGAGGAA..... 1422
 ::::: ::::: ::::: :::::
 461 IeGlnAlaSerGlyIaGlyValValGlnLeuSerGlyThrIleAlaAla 477
 1423AACCAAGGCTCGGTC...AGCGTGGCGCAAGGTAA 1454
 ::::: ::::: ::::: :::::
 478 GluLeuArgLeuGlyasnAlaGlySerValPheLysLeuAlaaspGlyTh 494

1010 nAlaThrIleAsnAspGlyValThrValThrThrGlyIleAlaGlyI 1027
2878GCCAAATTGAACGCTCAGGGAC 2901
1027 IeGlyPheAspGlyLysIleThrLeuGlySerValAsnGlyAsnGlyAsn 1043
2902 TTCGCGTTTATGTCGGAACCTCTCGGCTACCGCAGCGCAAAATTGAAGCT 2951
1044 ValArgPheValAspGlyIleLeuSerHisSerThrSerMetIleGlyI 1060
2952 GCGGGAAGTTCCGAAGGCACCTTACCTGGCT.....GTCA 2989
1060 rThrLysAlaAsnAsnGlyThrValThrThrLeuGlyAsnAlaPheValG 1077
2990 ACAATACCGCAGCAAGAA.....CCGTAAGTCTCAGCAATTCAGCGTA 3033
1077 LysAsnIleGlyAspSerAspThrProValAlaSerValArgPheThrGly 1093
3034 GTGGAAAGAAAGAACACACACCGCTGTCCGAAATCTTAATTTCACCT 3083
1094 SerAspGly.....GlyAlaGlyLeuGlnGlyLysnI 1104
3084 GCAAACGAAACACGTGATGCGCGCCATCGGCTTATCAGCTTATCCGA 3133
1104 eTyTrSerGlnValIleAspPheGlyThrTyTrAsnLeuGlyIleSerAsn 1120
3134 AAGACGCGCATGTCCTGCAATATCCGTCAAGACACAGAGCTTTC 3183
1121Ser 1121
3184 GACAACTCGCGCAGCGCGGAGAAACAGAGCCCTTGACGCGCAAAACA 3233
1122 AsnValIleLeuGlyGlyIleThrThr.....AlaIleAsnGly 1134
3234 GGCACAACTTGCGCCCAACACAGCGGCAAAAGACAGCGCAAGCC 3283
1135LysIleAsnLeuArgThrA 1141
3284 TTGACGCGCTGATGCGCGCGCGCAATCCACCGCAAAAGCAGAAAGT 3333
1141 snThrLeuThrPheAlaSerGlyThrSerThrThrGlyAsnAsnThrSer 1157
3334 GTTCCCGAAGCGCGCGCGCAGCAGCGGGAATGCCGCGCATTCGA 3383
1158 ILeGlyThrThrLeuThrLeuAlaAsnGlyAsnIleGlyAsnIleValI 1174
3384 GCGGAGAGAGAGAAAAACGGGTGACAGCGGATTAAGACACCGCTGG 3433
1174 eLeuGlnGlyAlaGlnValAlaSerThrThrThrGlyThrThrIleI 1191
3434 CGAAACAGCGCGAGCGGAAACCGG.....CCGCGTACACG 3471
1191 yValGlnAspAsnAlaAsnAlaAsnPheSerGlyThrGlnThrTyTrThr 1207
3472 GCGTTCGCCCGCGCGCGCGCGCGCGGATTCGCGCAACGCGACCC 3521
1208 LeuIleGlnGlyGlyAlaArgPheAsnGlyThrLeuGlyGlyPro 1222
3522 CCAAACGCAACCCCAACCGCAGCGGACCTGATC.....AGCC 3559
1223AsnPheValValThrGlySerAsnA 1231
3560 GTTATGCCAATAGCGTTTGTAGTGAATTTCCGCGCAGCTCAGACGCTT 3609
1231 rGpPheValAsnTyTrGlyLeuIleArg...AlaAlaAsnGlnAspTyTrVal 1246
3610 TTCCGCGTACAGAGCAATTCGACCGCTGTTCGCCGAAGCCGCCGAA 3659
1247 IleThrAlaGlyThrAsnAlaGlnValValThrAsnAspIleAlaAs 1263
3660 CGCCGTTTGG.....ACAAGCGGCA 3679
1263 nSerSerPheGlyGlyAlaProGlyValGlyGlnAsnValThrThrPheV 1280
3680 TCCGGGACACCAACACTACCGCTTCGAAG.....ATTCCGCGCC 3720
1280 aIAsnAlaThrAsnThrAlaIleAlaTyTrAsnAsnLeuLeuAlaLysAs 1296
3721 TACCCGCAACAAACCGACCTGCGCAATTCGGTATGCAGAAACCTCGG 3770
1296 nSerIleAsnSerIleAsnPheValGlyAlaIleValThrAspThrSerA 1313
3771 CAGCG..... 3775
1313 IaIleIleThrAsnAlaGlnLeuAspValAlaLysAspIleGlnAlaGln 1329
3776 ..GCGCGTGGGCACTCCTGTTTCGCAACCGGACCGGCAACACCTTC 3822
1330 LeuGlyAsnArgLeuGlyAlaLeuArgTyTrLeuGly.....ThrProG 1344
3823 GACGAGGCGCATCGCAACTCGGCAAGCTTCCACAGCTGCCCTTTCGG 3872
1344 uThrAlaGlnMetAlaGly.....ProGlnAlaGlyAlaIleP 1357
3873 GCATATACGCGCATCGGCAAGTTCGACATCGGATCGCGCGCGCGGTT 3922
1357 roAlaIleValAlaIleAlaGlyAspGlnAlaValaAspAsnValAlaTyTrGly 1373
3923 TTAGTACGCGCAGCCTTTCAGACG...GCATCAGAGCGCAAAATCGCGC 3969
1374 IleThrAlaLysProPheTyTrThrAspAlaHisGlnSerTyTrGlyI 1390
3970 GCGGTGCTGCATTCAGGATTCAGGCAAGATCGCGAGCTT.....T 4013
1390 y.....LeuAlaGlyTyTrLysAlaLysThrThrGlyAlaValIleG 1404
4014 GCGCGGATTCGCGCATCGAACG.....CACATCGCG 4044
1404 ILeuAspThrLeuAlaAsnAspAsnLeuMetIleGlyAlaIleAlaIleGly 1420
4045 GCAACGCGCTTTCGTCCAAAAAGCGGATTACGA.....TAGCA 4085
1421 IleThrLysThrAspIleLysHisGlnAspTyTrLysGlyAspLysTh 1437
4086 AAGCTCAATATCGCCACCCCGGCGCTTGCAATTCACCGCTACCGCGG 4134
1437 RasPheValAsn.....GlyPheSerPheSerLeuTyTrGlyAlaG 1450
4134 4134
1450 InGlnLeuValLysAsnPhePheAlaGlnGlySerAlaIlePheSerLeu 1466
4134 4134
1467 AsnGlnValLysAsnLysSerGlnArgTyTrPhePheAspAlaAsnGlyAs 1483
4134 4134
1483 nMetSerLysGlnIleAlaIleGlyHisTyTrAspAsnMetThrPheGlyG 1500
4135 ..GGCATTAAGCAGATTAATTCATCAACCGCGCGCAACATTCATTCATC 4182
1500 LysAsnLeuThrValGlyTyTrAspTyTrAsnAlaMetGlnGlyValLeuVal 1516
4183 ACGGCTTATTTGAGCGCTGCTGCTATACGATGCGGCTTCGCGCAAAAGTCG 4232
1517 ThrPrometAlaGlyLeuSerTyTrLeuLysSerSerAspGlnAsnTyTrLys 1533
4233 AAGCGCGTCAATATACCGCGCTATTCGCGCAGGATTCGCGCAAA...ACC 4279
1533 sGluThrGlyThrThrValAlaAsnLysGlnValAsnSerLysPheSerA 1550
4280 GCACTGCGGAGATGGCGGCTTAACCGCGCAATCAAGGTTTCACGCTGTCC 4329
1550 sPArgThrAspLeuIleValGlyAlaLysValAlaGlySerThrMetAsn 1566

4330 CTCACGCTGCCGCCGCAAGGCGCAATGTGACG 4368
 1567 ILeHraspleuAlaValTyr...ProGluValHisAla 1578
 seq_name: Swissprot_40: YDEK_ECOLI

seq_documentation_block:
 ID YDEK_ECOLI STANDARD; PRT; 1325 AA.
 AC P32051; P76140; P77168;
 DT 01-OCT-1993 (Rel. 27, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hypothetical lipoprotein ydek precursor (ORF1).
 GN YDEK OR ORF1 OR B1510.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia.
 OX NCBI_TaxId=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX STRAIN=K12 / MG1655;
 RX MEDLINE=97426617; PubMed=9278503;
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.,
 RT "The complete genome sequence of Escherichia coli K-12.";
 RL Science 277:1453-1474(1997).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX STRAIN=K12;
 RX MEDLINE=97251357; PubMed=9097039;
 RA Alba H., Bada T., Fujita K., Hayashi K., Inada T., Isono K., Itoh T.,
 RA Kasai H., Kashimoto K., Kimura S., Kitakawa M., Kitagawa M.,
 RA Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,
 RA Nakade S., Nakamura Y., Nashimoto H., Nishio Y., Oshima T., Saito N.,
 RA Sempel G., Seki Y., Sivasubraman S., Tagami H., Takeda J.,
 RA Takemoto K., Takeuchi Y., Wada C., Yamamoto Y., Horinouchi T.,
 RT "A 570-kb DNA sequence of the Escherichia coli K-12 genome
 corresponding to the 28.0-40.1 min region on the linkage map.";
 RL DNA Res. 3:363-377(1996).
 RN [3]
 RP SEQUENCE OF 595-1325 FROM N.A.
 RX MEDLINE=94100243; PubMed=8274505;
 RA Catwright P.J., Timms M.W., Litgow T., Hoef P.B., Hoogenraad N.J.,
 RT "An Escherichia coli gene showing a potential ancestral relationship
 to the genes for the mitochondrial import site proteins ISP42 and
 MOM38.";
 RL Biochim. Biophys. Acta 1153:345-347(1993).
 CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a lipid anchor
 (Potential).
 CC -1- SIMILARITY: TO E. COLI YFAL.
 CC -1- SIMILARITY: SOME, TO FUNGAL MITOCHONDRIAL IMPORT SITE PROTEINS
 ISP42 AND MOM38.
 CC -1- CAUTION: REF.3 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO A
 FRAMESHIFT IN POSITION 653.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AE000248; AAC74583.1; -
 DR EMBL; D90793; BA15190.1; ALT_INIT.
 DR EMBL; D90794; BA15197.1; ALT_INIT.
 DR EMBL; X73295; CA51730.1; ALT_FRAME.
 DR PIR; S34315; S34315.
 DR ECGene; EG11780; ydek
 DR PROSITE; PS00013; PROKAR_LIPOPROTEIN; 1.

KW Hypothetical protein; Membrane; Lipoprotein; Signal;
 KW Complete proteome.
 FT SIGNAL 1 18 POTENTIAL.
 FT CHAIN 19 1325 . HYPOTHEICAL LIPOPROTEIN YDEK.
 FT LIPID 19 19 N-ACYL DIGLYCERIDE (POTENTIAL).
 FT CONFLICT 884 884 N -> K (IN REF. 3).
 FT CONFLICT 1317 1317 M -> S (IN REF. 3).
 FT SEQUENCE 1325 AA; 136514 MW; 26A3A066FA19AD7D CRC64;
 alignment_scores:
 Quality: 235.00 Length: 1487
 Ratio: 0.338 Gaps: 83
 Percent Similarity: 46.738 Percent Identity: 20.108
 alignment_block:
 US-09-303-518d-653 x YDEK_ECOLI ..
 Align seg 1/1 to: YDEK_ECOLI from: 1 to: 1325
 613 CGGTGATGAGAGAGACGACCAATACCGCAAGTTCATCATATGTC 662
 ||||| : : : : : ||| : : : : :
 25 ATGATGAGTCTGATGATGATGATGATGATGATGATGATGATGAT 41
 ||||| : : : : : ||| : : : : :
 663 AAGCGCATATTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 712
 :
 41 rthlypserfargleuthrleuvalleuvalleuvalleuvalleu 58
 ||||| : : : : : ||| : : : : :
 713 CA...GCTGTGACAGTCACTTATGATGATGATGATGATGATGAT 759
 || :
 58 eralaserglyalaserleuvalleuvalleuvalleuvalleuval 73
 ||||| : : : : : ||| : : : : :
 760 CCATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 809
 :
 74leasphrasp..... 77
 ||||| : : : : : ||| : : : : :
 810 AATGTTATGATGATGATGATGATGATGATGATGATGATGATGAT 859
 ||||| : : : : : ||| : : : : :
 110 Valilleglyalaserleuvalleuvalleuvalleuvalleuval 126
 ||||| : : : : : ||| : : : : :
 915 AGATGCTG...TTCATGATGATGATGATGATGATGATGATGATG 943
 ||||| : : : : : ||| : : : : :
 126 ythtrpargleuthrleuvalleuvalleuvalleuvalleuval 143
 ||||| : : : : : ||| : : : : :
 860 TGCACACA.....GGCAC.....CCG 876
 ||||| : : : : : ||| : : : : :
 93 eusnilleuvalleuvalleuvalleuvalleuvalleuvalleu 109
 ||||| : : : : : ||| : : : : :
 877 TATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 914
 ||||| : : : : : ||| : : : : :
 110 Valilleglyalaserleuvalleuvalleuvalleuvalleuval 126
 ||||| : : : : : ||| : : : : :
 915 AGATGCTG...TTCATGATGATGATGATGATGATGATGATGATG 943
 ||||| : : : : : ||| : : : : :
 126 ythtrpargleuthrleuvalleuvalleuvalleuvalleuval 143
 ||||| : : : : : ||| : : : : :
 944 GAGAT.....ACCATTCAGTATTCATGATGATGATGATGATGAT 981
 ||||| : : : : : ||| : : : : :
 143 lylinseryglythrlthleuvalleuvalleuvalleuvalleu 159
 ||||| : : : : : ||| : : : : :
 982 GCGAATACGTTT...TTTACGATATATATATATATATATATAT 1028
 ||||| : : : : : ||| : : : : :
 160 Glyllyglyleuvalleuvalleuvalleuvalleuvalleuval 176
 ||||| : : : : : ||| : : : : :
 1029 TGCACACAATGATGATGATGATGATGATGATGATGATGATGAT 1078
 ||||| : : : : : ||| : : : : :
 176 nvalilleglylinseryvalleu.....Thrt 186
 ||||| : : : : : ||| : : : : :
 1079 TTCATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1110
 ||||| : : : : : ||| : : : : :
 186 hrglueuphegluileglyserlyglythrlthleuvalleuval 202
 ||||| : : : : : ||| : : : : :
 1111 GCAAGACAGCTGTTTATCATGCTGCGAGGTGGGTCAAGATGATGAC 1160
 ||||| : : : : : ||| : : : : :
 203 Asplyslyglythrlthleuvalleuvalleuvalleuvalleu 219
 ||||| : : : : : ||| : : : : :
 1161 CAGACTGATATGATGATGATGATGATGATGATGATGATGATGAT 1210

219 agly...Serasnlglylnvalval.....Valglulysgly...glyg 232
221 AATGATACCTTACCGACACATCATACCAAGCGCGCGGCTTGTATTT 1260
232 lurtplleuilelysasnln.....AspSerSerilegluphe 244
1261 GAG...GTAATTTTACGGCTCGCTTAAACCAAGCAACGTGGCAGG 1307
245 Glulileglyasn.....GlnGl 250
1308 CGCGGCG...GTTCAATATCAGTATGCGACCTTACTTGGAAA.... 1350
250 ythnglyglualathrllearglulglyleuvalThrlaglulnsnt 267
1351GTAACGGCGTGGCAACGACCGCTGTCCAAAATCGGCAAGC 1395
267 hrlleileglyllyasnlatthr.....Glyilegly 277
1396 ACCTGCTGCTTCAAGCCAAAGGGAACCAAGCTCGCTGACGCG... 1443
278 Thrleuasnvalgln.....AspGlnAspSerValilethrya 290
1444GGCAGCGTAAAGTCATCTAGATC 1468
290 lartgargleuTyraSnlglyrphleglyasnlglyThrlvalasnilesera 307
1469 AGCAGGCGACGATCAAGCAAAAAAACGCCCTTAGTGAAATCGGCTTG 1518
307 snasnlglyleuileasnlnlysgln.....TyrSerleuvalGlyval 321
1519 GTCAGCGGC...AGGGGAGCGGTGCAACTGAATGCCGATATCAGTTCA 1565
322 GlnAspGlySerHisgllyalValasnvalThr..... 332
1566 CCGCGCAAAACTCTATTGGCTTCCGCGCGGACCTTGGATTGACG 1615
333AspLysglYHisTrpAsnphelugly.....Thrg 343
1616 GGCATTCGCTTGTCCACCGCATCAAAATACCGATGAGGGGGGAG 1665
343 lylulalphearglyrileTyrlleGlyAspAlaGlyaspGlyluleu 359
1666 ATTGTCAACCAATCAGACAAAGAAATCCAGCTTACCATTAACAGCA 1715
360 AsnvalSerSerGlnGlyllyvalAspSerGlylleThrlAlaGlyme 376
1716 TAAGATATTACTACACCGGCAATTAACAACATTGGATGCAAAAAA 1764
376 llysglu...ThnglyThrglyasnleThryalLysAspLysAsnServ 392
1765GAAATTCCTTACAGCGTGTGGCGAG 1794
392 alileThrasnleuglyThrasnleuglyTyraSpely..... 404
1795 AAAGATGCAACCAAAAGACGCGGCTCAATTCGATTAACACCGGA 1844
405HisglYlumeTasnleIleSerasnln..... 413
1845 AGAAGCGATCGACTTACTGCTTCCGGGAGAACAAATTTAAAC.... 1890
414GlyleuvalValSerasnlglyllySerleuglyTyrg 427
1891GGCAATATCAGCAAAACAAAGCGAAATCTTTTC 1926
427 lylulThrglyalGlyasnvalSerileThrglyllymetrip... 442
1927 ACGGAGACCGACCGGCTTACATCAATTAAGCAAGCGGTGCTC 1976
443GlnvalAsnLysasnvalTyThrlleGly..... 453
1977 AAAAAAGAGTATCCACAGAGAAATCGTGTGGCAACGATTGGA 2026
::: ||:::
454ValAlaGlyVal..... 457
2027 TCGACCGCATTTAAAGCGGAACCTTCATATTACGCGGACAGACG 2076
458GlyAsnleuAsnileSerAspLyllylys 467
2077 GTGGTTTCCCGCAATGTTCCAAAGTGGAGCGCATTTGCAATTAAAGCA 2126
468 phevalSerGlnAsnleThrphleu...Glyasp..... 478
2127 TCACGCCCAAGCAGTTTCGTGTGCGACCGCATCAAGCCACACATCT 2176
478 478
2177 GTACACGTTGCGATGAGCGGCTGTGACAGTTTACGCAAAACCAT 2226
479LysAlaSerGlyileGlyThrlleuAsnleuMetAspAlaThrSer 493
2227 ACCGACGATTAAGTATGCTTCAATTGAGCAAGCCGACATCAGAGCGAA 2276
494 SerPheAspThrValGlyllyleasnvalGlyasnphely...SerGlyyl 509
2277 TGTACGCTTGGCGATCAAGCTCATTTAAATCTCAGAGCTTGGCACAC 2326
509 eValasnvalSerasnlglyAlaThrlleuasnSerThrglyllyrGlyPheI 526
2327 TCACGCGCATTTAGTCAGGCGGA.....GACACG 2358
526 leGlyllyAsnAlaSerGlyllyGlyleValasnileSerThraspSer 542
2359 CACTATACGTTACCGGACGCGCAC.....CAAAACG 2393
543 leutrpasnleuLyfThrSerSerThrasnAlaGlnleuGlnvalGl 559
2394 C.....AACCTCAGCTGTGGCAATGCCACG 2422
559 yAlleuglyThrglyluleuasnleThrThnglyllyleVallysa 576
2423 CACATTTAATCAAGCCCATTTAAACGGCAACATCGGCT..... 2463
576 lartgAspThrglnleleuasnAspLysSerlyGlyAspvalArg 592
2464TCGCAATGCT.....TCATTTATCTAACCAACACG 2498
593 ValAspGlyGlnAsnSerleuGlnThrPheAsnMetTyraGlyl 609
2499 CGTACAAACGCGCAGTGAAGCTTTCGACACAGCTTAAGGCAACGTA 2547
609 rSerGlyThrglyThrlleuThrlAsnAsnlglyThrlleuasnvalG 626
2547 2547
626 lulglyGlylvalTyrlleuGlyvalPheGlnProlaValGlyThrlleu 642
2548ACCAT.....TCGCACTCAAGCGCAATGCTC 2576
643 AsnleleGlyAlaAlaHisGlylulAlaAlaAspAlaGlyPheileTh 659
2577 CCTACCGCATTAAGCAGTATTC..... 2598
659 rAsnAlaThryllyvalGlnphelyleuglylulglyvalPhevalPhea 676
2599CATTTGAACACAGC.....CGCTTACC 2622
676 snHlstrAsnAsnSerAspAlaGlyTyrglnvalAspMetleuileThr 692
2623 GGAATAATCAGCGCGGCAAGGATACGCAATTAACATAAGACAGCGA 2672
693 GlyAspAspLysAspGlyLys.....ValIleHis.....AspAlaGl 705
2673 ATGACGCTG.....CGTGGGACGCGAATTAAGCAATTAACCTTG 2716
705 yHlstrThrPheasnAlaGlyasnThryllySerGlyLysThrlleuVala 722

```

2717 ACAAGCCACCATTTACTCAATTCGGCTATCGACACGATCGCGACG 2766
      :::::::::::::::::::: ||| ||| |||
722 snasplyleuethrillealaser.....HisThrAlaasplly 735
      :::::::::::::::::::: ||| ||| |||
2767 GCGCAACCGCGAGTCGCGACGATCGCGCGCGCGCTGCGCGCGTTC 2816
      :::::::::::::::::::: ||| ||| |||
736 Val...ThrGlyMetClyser.....Se 742
      :::::::::::::::::::: ||| ||| |||
2817 CCTATTATCCGTTTACCGCGCAACTTGGCGAGATCCGTTTCAACACGC 2866
      :::::::::::::::::::: ||| ||| |||
742 rgluValThrIleAlaasPro.....GlyThrIL 752
      :::::::::::::::::::: ||| ||| |||
2867 TCAGCGTAAACGGCAATTAAGCTGACGGACATCCGCTTTATGTCG 2916
      :::::::::::::::::::: ||| ||| |||
752 euasplyleuAlaserThrAsnSerAlaGlyAspTyrThrLeuThrAsn 768
      :::::::::::::::::::: ||| ||| |||
2917 GAACCTCTTCGCTACCGCGCGCAAAATTGAAGTGGCGGAAGTTCCGA 2966
      :::::::::::::::::::: ||| ||| |||
769 AlaLeuLysGly.....AspLysLeuMetArgValGlnLeuSerSerSe 783
      :::::::::::::::::::: ||| ||| |||
2967 AGGCACTTACACCTTGGCTGTCACAAATACCGCGCAACCGCGTAATGC 3016
      :::::::::::::::::::: ||| ||| |||
783 rAspLysMetPheGlyPheThrHisAlaThrGlyThrGluPheAlaGly 800
      :::::::::::::::::::: ||| ||| |||
3017 TCGAGCAATTGACGGTAGTGAAGGAAGAACACACACCGCTGCCGA 3066
      :::::::::::::::::::: ||| ||| |||
800 AlaAlaGlnLeu.....LysAspSerThr..... 807
      :::::::::::::::::::: ||| ||| |||
3067 AATCTTAATTTCACCTTCGCAAAACGACACGTCGATGCCGCGCATGGCG 3116
      :::::::::::::::::::: ||| ||| |||
808 .....PheThrLeuGlnArgAsp.....AsnThrAlaAlaLeuTh 819
      :::::::::::::::::::: ||| ||| |||
3117 TTAATGCTTATCCGCAAGACGGGAGTTCCGCTGCATATCCGGTCA 3166
      :::::::::::::::::::: ||| ||| |||
819 rHisAlaMetLeuGlnSerAspSerIleAsnThrThrSerValLysValG 836
      :::::::::::::::::::: ||| ||| |||
3167 AAGAACAAGAGCTTTCGCAAAACTCGGACGAGGAGGAAGAACAGAGCC 3216
      :::::::::::::::::::: ||| ||| |||
836 LysGlnIleSerIleGlyLysLeuAlaMetAsnGlyLysThrIlePhe 852
      :::::::::::::::::::: ||| ||| |||
3217 GCCTTACGCGCAAAACAGGACACACTTGGCGCGCAACACAGCGCGAAA 3266
      :::::::::::::::::::: ||| ||| |||
853 AspThrAspIleProAlaAlaThrIleAla..... 862
      :::::::::::::::::::: ||| ||| |||
3267 AGCAACACGGCGCAAGCTTGAAGCGCTGATGGCGCGGCGCAATGCCA 3316
      :::::::::::::::::::: ||| ||| |||
863 GlnGlyTyrIleSerValAspThrLeuValAlaGlyAlaGlyAspTyr 879
      :::::::::::::::::::: ||| ||| |||
3317 CCGAAAGGCGAGAAAGT..... 3333
      :::::::::::::::::::: ||| ||| |||
879 hTrpTrpLysGlyArgAsnTyrGlnValAsnGlyThrGlyAspValLeuIle 895
      :::::::::::::::::::: ||| ||| |||
3334 ...GTTGCGGAACCG..... 3345
      :::::::::::::::::::: ||| ||| |||
896 AspValProLysProTrrAsnAspPrometAlaAsnAsnProLeuThrTh 912
      :::::::::::::::::::: ||| ||| |||
3345 ..... 3345
      :::::::::::::::::::: ||| ||| |||
912 rLeuAsnLeuLeuGlnHisAspSerHisValGlyValGlnLeuValL 929
      :::::::::::::::::::: ||| ||| |||
3346 ...GCCGCGACAGCGCGGGAAGATCCGCGCATTAATGACAGCGGAGAA 3393
      :::::::::::::::::::: ||| ||| |||
929 ysaLacGlnThrValIleGlySerGlyLysLeuThrLeuArgAspLeu 945
      :::::::::::::::::::: ||| ||| |||
3394 GAGAAAAACGGGTGACAGCGCGATAAAGACCGCGCTTGGCGCAACAGCG 3443
      :::::::::::::::::::: ||| ||| |||
946 GlnGlyAspGlyValAlaGlnAlaAspLysThrLeuHisIleAlaGlnAsnG 962
      :::::::::::::::::::: ||| ||| |||
3444 CGAAGCGGAACCGCG.....CCGCGTACCAACCGCGCTTGC 3478
      :::::::::::::::::::: ||| ||| |||
962 yThrValValAlaGlnGlyAspTyrGlyPheArgLeuThrThrAla.... 977
      :::::::::::::::::::: ||| ||| |||

```

```

3479 CCGCGCGCGCGCGCGCGCGCGGATTTGCCGCAACGCGAGCCCAACGG 3528
      :::::::::::::::::::: ||| ||| |||
977 ..... 977
      :::::::::::::::::::: ||| ||| |||
3529 CAACCCCAACCGACGCGCGACCTGATCAGCGTTATGCCAATAGCGTTT 3578
      :::::::::::::::::::: ||| ||| |||
978 .....ProGlyAsnGlyLeu.....TyrValAsnTyrGlyLe 988
      :::::::::::::::::::: ||| ||| |||
3579 G..... 3579
      :::::::::::::::::::: ||| ||| |||
988 uLysAlaLeuAsnIleHisGlyGlnLysLeuThrLeuAlaGlnHisG 1005
      :::::::::::::::::::: ||| ||| |||
3580 ...AGTGAATTTCCGCGCGCGCTCAGACGCGTTTCCGCTACGAGCA 3627
      :::::::::::::::::::: ||| ||| |||
1005 LysGlyAlaTyrGlnAlaThrAlaAspMetSerAlaLysIleGlyGln 1021
      :::::::::::::::::::: ||| ||| |||
3628 TTGACCGCGCTTGGCGGAGACCGCGCGACCGCGTTTGGCAAGCG 3677
      :::::::::::::::::::: ||| ||| |||
1022 GlyAspLeuAlaIleAsnThrValArgGlnValSerLeu...SerAsnG 1037
      :::::::::::::::::::: ||| ||| |||
3678 CATCGCGGACACCAACACTACCGTTGCGCAAGATTTCCGCGCTTACCGC 3727
      :::::::::::::::::::: ||| ||| |||
1037 yGlnAsnAsp.....TyrGlnGly.....AlaThrTyrValG 1048
      :::::::::::::::::::: ||| ||| |||
3728 AACAAACCGACCTCGCG.....CAATCGGTATGCAAGAA 3762
      :::::::::::::::::::: ||| ||| |||
1048 ImetGlyThrLeuArgThrAspAlaAspLysAlaLeuGlnAsnThrArg 1064
      :::::::::::::::::::: ||| ||| |||
3763 AACCTCGGCGAGCGCGCGCGCGCATC.....CTGTTTGCAGCAACG 3806
      :::::::::::::::::::: ||| ||| |||
1065 GlnLeuAsnIleSerAsnAlaIleAlaIleValAspLeuAsnLysThrG 1081
      :::::::::::::::::::: ||| ||| |||
3807 GACCGGAACACCTTTCGACGACGCGCATCGCAACTCGCGCGCTTCCC 3856
      :::::::::::::::::::: ||| ||| |||
1081 nThrValGlnThrPheThrGlyGlnMetGlySerThrValLeuPheLysG 1098
      :::::::::::::::::::: ||| ||| |||
3857 AGCGTCCGTTTTCGCGCAATACGCGCATCGCGAGTTTCGACATCGGATC 3906
      :::::::::::::::::::: ||| ||| |||
1098 LysGlyAlaLeuThrValAlaAsnLysGly..... 1106
      :::::::::::::::::::: ||| ||| |||
3907 AGCGCGGCGCGGTTTATAGCGAGCGCTTTCAGACGCGCATAGAGG 3956
      :::::::::::::::::::: ||| ||| |||
1107 .....GlyIleSerGlnGlyLysLeuThrGlyLysGlnLe 1119
      :::::::::::::::::::: ||| ||| |||
3957 CAATTCGCGCGCGCGCGCTGCTCATTTAC...GGCATTCAGCAAGATACC 4003
      :::::::::::::::::::: ||| ||| |||
1119 uAsnValThrGlyLysThrLeuAlaIleGlnGlyLeuAsnAlaArgTyr 1136
      :::::::::::::::::::: ||| ||| |||
4004 GCGCAGGTTTGGCGGATTCGACATCGAACCGCGCATCG..... 4042
      :::::::::::::::::::: ||| ||| |||
1136 snAla...LeuThrSerIleSerProAsnAlaGlnValSerLeuAspAsnTh 1152
      :::::::::::::::::::: ||| ||| |||
4043 .....GCGCAACGCGCTATTTGCTGCATAA 4067
      :::::::::::::::::::: ||| ||| |||
1152 rGlnGlyLeuGlyArgGlyAsnIleAlaAsnAspGlyLeuLeuThrLeu 1169
      :::::::::::::::::::: ||| ||| |||
4068 AGCGGATTTAC...GATPAGAAACGTCATATGCGCACCGCGGCGCTTG 4114
      :::::::::::::::::::: ||| ||| |||
1169 ysaAsnValThrGlyLysLeuArgAsnSerIleSerGlyLysGlyIleVal 1185
      :::::::::::::::::::: ||| ||| |||
4115 CATTCACCGCT.....ACCGCGCGGCG 4137
      :::::::::::::::::::: ||| ||| |||
1186 SerAlaThrAlaArgThrAspValGlyLeuAspGlyLysAsnSerArgPh 1202
      :::::::::::::::::::: ||| ||| |||
4138 APTAAGGAGATTTATTCATCAACCGCGCAACACATTTCCATCAGCGC 4187
      :::::::::::::::::::: ||| ||| |||
1202 eVal...GlyGlnPheAsnIleAspThrGlySerAla...LeuSerValAsnG 1218
      :::::::::::::::::::: ||| ||| |||
4188 TTATTGAGCTGCTTACGAGTGGCGCTTCCGGAAGTCCGAAAGCG 4237
      :::::::::::::::::::: ||| ||| |||
1218 uGlnLysAsnLeuGly.....AspAlaSerVal..... 1227
      :::::::::::::::::::: ||| ||| |||
4238 GCGTCATATCCGCGCTATTTGGCGCAGGATTTCCGCAAAACCGCGAGTGC 4287
      :::::::::::::::::::: ||| ||| |||

```



```

1228 ..IIeAsnAsnGlyLeuLeu.....ThrlIleSerThr 1237
4288 GAA.....TGGGCGGTAAACCGCAATCAAGCT..... 4317
1238 GluArgSerTrpAlaMetThrHisSerIleSerGlySerValTh 1254
4318 .....TTACAGCTGTCCCTCCAGCGTCCGCCGCCCA 4348
1254 rlyLeuGlyThrGlyIleThrLeuThrLeuAsnAsnSerAlaIleTyrg 1271
4349 AGGGG 4353
1271 Ingly 1272

```

seq_name: SwissProt_40:PM21_CHLPP

seq_documentation_block:

ID PM21_CHLPP STANDARD; PRT; 1609 AA.

AC 092605; Q9RB58;

DT 16-OCT-2001 (Rel. 40; Created)

DT 16-OCT-2001 (Rel. 40; Last sequence update)

DE Probable outer membrane protein pmp21 precursor (Polymorphic membrane protein 21).

GN Pmp21 OR CPN0963 OR CP0897.

OS Chlamydia pneumoniae (Chlamydia pneumoniae).

OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.

OX NCBI_TaxID=83558;

NP [1]

RP SEQUENCE FROM N.A.

RC STRAIN=CM1029;

RA MEDLINE=920606; PubMed=10192388;

RA Kalman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,

RA Olinger L., Grimwood J., Davis R.W., Stephens R.S.,

RT "Comparative genomes of Chlamydia pneumoniae and C. trachomatis";

RL Nat. Genet. 21:385-389(1999).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=AR39;

RA MEDLINE=20150255; PubMed=10684935;

RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,

RA White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,

RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,

RA Swinn M., Nelson W., Deboy R., Kolonay J., McClarty G., Salzberg S.L.,

RT "Genome sequences of Chlamydia trachomatis Mopn and Chlamydia

trachomatis AR39."

RL Nucleic Acids Res. 28:1397-1406(2000).

RN [3]

RP SEQUENCE FROM N.A.

RC STRAIN=J138;

RA MEDLINE=20330349; PubMed=10871362;

RA Shira T., Ishii K., Hattori M., Tabuchi M., Kishi F., Ouchi K.,

RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138

from Japan and CM1029 from USA."

RL Nucleic Acids Res. 28:2311-2314(2000).

RN [1]

RP SUBCELLULAR LOCATION: CELL WALL SURFACE (ELEMENTARY BODIES)

(POTENTIAL).

- SIMILARITY: BELONGS TO THE PMP OUTER MEMBRANE PROTEIN FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb.ch/announce/>

CC or send an email to license@isb.ch).

CC -----

CC EMBL; AE001676; AAD19099.1; -

CC EMBL; AE002248; AAF38684.1; -

CC EMBL; AP002548; BAA9171.1; -

```

DR PHC1-2DPAGE; Q9RB58; -
DR TIGR; CP0897; -
DR InterPro; IPR003357; OMP.
DR Pfam; PF02385; OMP; 1.
KW Outer membrane; Signal; Multigene family; Complete proteome.
FT SIGNAL 1 30
FT CHAIN 31 1609
FT CONFLICT 420 420
SQ SEQUENCE 1609 AA; 170865 MW; 2604C3E9FC4024CB CRC64;

```

alignment_scores:

Quality	Ratio	Length	Gaps
224.00	0.367	1427	76
Percent Similarity: 42.747	Percent Identity: 19.131		

alignment_block:

US-09-303-518D-653 x PM21_CHLPP ..

Align seg 1/1 to: PM21_CHLPP from: 1 to: 1609

```

1108 ACAGCAAGA.....GAACCTGTTATCATGCTGCGAGGTGGGTACAG 1151
|||||
324 ThrAlaArgAsnHisGluValGlnAsnSerAlaGlyIleLeuSe 340
1152 TTATGACCCACAGCTGAATATGAGAAATTTCTTTATTCACAAAG 1201
|
340 r.....ValValGly 344
1202 GAAAGTGATGATTAATCTTACCAGCAC..ATCACCAAGCGCGGCG 1248
|||||
344 snlySGlValAlaIleValGluIlyAsnSerAlaGluIlySerSnlly 360
1249 GGTGTGATTTTGAAGGATTTTACGGTCTCGCTCAAAACCAAGAAC 1298
|||||
361 GlyAlaPheAlaCysGlySerPheValTySer.....AsnAsnGlu.. 374
1299 GTGGCAAGCGCGCGCTTCATATGATGATGATGATGATGATGATG 1348
|||
375 .....AsnThrAlaLeuTrp 380
1349 AAGTAAAC.....GGCGTGCAAAAGACCGCGCTCCCAAAATC 1386
|||
380 ysglAsnGlnAlaLeuSerGlyAlaIleSerSerAlaSerAspIle 396
1387 GGC...AAGGCACGCTCTGCTGTCACCAAGGGGAAACCAAGGCTC 1433
|||||
397 AspIleGlnGlyAsnCySerAlaIleGluPheSerGlyAsnGlnSerle 413
1434 GTTCAGCGTGCGGCGAGTAAAGTCATCTTAGATCAGCAGCGGACATC 1483
|||||
413 wleAlaLeuGlyGlnHisIleGlyLeuThrAsp..... 424
1484 AAGCAAAACCAAGCCTTTAGTAATCGCTTGTCAGCGGCGAGGGG 1533
|||
425 .....PheValGlyGlyAlaLeuAlaIleGlnGly 435
1534 ACGGTGCACTG.....AATGCCGATATCAGTTC.....AACCCGA 1571
|||||
436 ThrLeuThrLeuArgAsnAsnAlaValGlnCyValIlyAsnTrpSe 452
1572 CAACCTATTTTCGGC...TTCCGGCGGCGAGCTTGATTTGAACGGGC 1618
|||||
452 rlystThrHisGlyAlaIleLeuAlaGlyThrValAspLeuAsn.... 467
1619 ATTGGCTTTCTTCACCGCATTTAAATTCGATGAGGCGCGATGATT 1668
|||||
468 .....GluThrIleSerGluValAlaPhe... 475
1669 GTCAACCAATCAAGCAAAAGATTCACCGTTACCTTACAGGACATAA 1718
|||||
476 .....LysGlnAsnThrAlaIleLeuThrGlyIlyAl 486

```



```

1043 rAlaProSerLeuIlePheAsnSerLysProGlnAspSerAlaGln 1060
2888 ACGGTGAGGAGACATTCCGCTT.....ATGTGGAA 2919
:::|||||:::|||||
1060 IShISgluGlyThrIleArgPheSerArgIValSerLysIleProGln 1076
:::|||||:::|||||
2920 CTCTTGGCGCTACCGCAGCGGCAAAATTGAACTGGCGAAATTCCGAGG 2969
:::|||||:::|||||
1077 ILeAlaIleIleGlnGlyThrLeuAlaLeuSerGlnAsnIleGlu 1093
:::|||||:::|||||
2970 CACTTACACCTTGCTGTCAACAATACCGCAACGAAACCGTA..... 3012
1093 uTrPLeuAlaGlyLeuLysGlnIleThrLysSerIleValLeuSera 1110
3013 .....AGTCTGAGCAATTGACGCTAGTGAAGAAAGCAACACACCG 3057
|||||:::|||||:::|||||
1110 laGlySerIleLeuArgIlePheAspSerGlnValAspSerAlaPro 1126
3058 CTG...TCGAAATCTTAATTTCACCTGCAAAACGAAACAGCTGATCG 3104
|||||:::|||||:::|||||
1127 LeuProThrGlnAsnLysGlnIleThrLeu.....ValSerAl 1139
3105 CGGCGCATGGCGCTTATCAGCTTATCCGCAAAAGCGGAGTCCGCTCG 3154
|||||:::|||||:::|||||
1139 aGlyVal.....GlnIleAsnMetS 1146
3155 ATAATCGGTCAAAGAACAGACCTTCCGACAACTCGCAAGCGCGGA 3204
:::|||||:::|||||
1146 eSerProThrProAsnLysAsp.....LysAlaVal 1156
3205 GAAACGAGAGCC.....GCTTGAGCGGCAAAACAGGACACA 3239
|||||:::|||||:::|||||
1157 AspThrProValLeuAlaAspIleIleSerIleThrValAspLeuSer 1173
3240 ACTTGCGCGCAAAACAAACAGCGGCAAAAGCAACAGCTT..... 3285
|||||:::|||||:::|||||
1173 rPheVal.....ProGlnIleAspIleThrLeuProLeuProP 1186
3286 ..GACGCGCTGATTGCGCGCGGCGCAATGCCAGCAAGGCA..... 3327
roGluIleIleIleIleProLysGlyThrLysLeuHisSerAsnAlaLeasp 1202
1186 roGluIleIleIleIleProLysGlyThrLysLeuHisSerAsnAlaLeasp 1202
3328 GAAAGTGTGGCGAACCGCGCGGCGAGCGGCGGAAATGCCGCGCT 3377
|||||:::|||||:::|||||
1203 LeuLysIleIleAspProThr...AsnValGlyTyrGluAsnHisAlaLe 1218
3378 TATCGAGGCGGAGAGAAAGAAAAACGGGTGCGAGCGGATAAAGACACG 3427
|||||:::|||||:::|||||
1218 uLeuSerSerHis.....LysAspIleP 1226
3428 CCTTGGCGAAACAGCGCGGCAAGCGGAAACCGCGCGCTAAC...ACGCGC 3474
|||||:::|||||:::|||||
1226 roLeuIleSerLeuLysThrIleGlnIleGlyMetThrGlyThrProThrAla 1242
3475 TTCGCCCGCGCGCGCGCGCGGATTTGGCGCAACCGCAGACCCCA 3524
|||||:::|||||:::|||||
1243 AspAlaSerLeuSerAsnIleLysIleAspValSerLeuProSerIleThr 1259
3525 ACCGCAACCCCAACCGCAGCGGAGCTGATCAACCGTTATGCCAATACG 3574
|||||:::|||||:::|||||
1259 rPro.....AlaThrTyrGlyHisThrG 1267
3575 GTTTCAGTGAATTTCCGCCCAACGCTCAACAGCGTTTCCGCTACAGAC 3624
|||||:::|||||:::|||||
1267 LysValTyrSerGluSerLys..... 1273
3625 GAATTGACCGCGTGTTCGCCGAGACCGCGCGC..... 3657
|||||:::|||||:::|||||
1274 .....MetGluAspGlyArgLeuValValGlyTyrPgl 1284
3658 .....A 3658
1284 nProThrGlyTyrLysLeuAsnProGluLysGlnGlyAlaLeuValLeuA 1301

```

```

3659 ACCCGCTTTGACACGCGCATCCGCGGACACCAACACTACCGTTCCGA 3708
||:::|||||:::|||||
1301 snAsnLeuThrSer.....HisTyr.....Thr 1308
3709 GATTTCGCGCGCTACCGCAACA.....ACCGA 3737
|||||:::|||||:::|||||
1309 AspLeuAlaGlyAlaLeuLysGlnIleIlePheAlaHisThrIleAlaG 1325
:::|||||:::|||||
3738 CCTGCGCAAAATCGGTATGCAAAAAACCTCGCGCGCGCGCGGCA 3787
:::|||||:::|||||
1325 nArgMetGluLeuAspPheSerThrAsnValThrPheLysGlyLeuGly 1342
3788 TCCTG.....TTTTCGCACAAAC 3804
|||||:::|||||:::|||||
1342 aIValGluAspCysGlnAsnIleGlyIlePheAspGlyPheLysHis 1358
3805 CGGACCGCAACACCTTCGACGAGCGCATCGCAACTCGCGAGCGCTTC 3854
|||||:::|||||:::|||||
1359 LeuThrGlyTyr.....AlaLeuGlyLeuAspThrGlnLeuVal 1371
3855 CCAC.....GGTCCGTTTTCGGGCAATACGCGATCGCAGGT 3892
:::|||||:::|||||
1371 GluAspPheLeuIleGlyGlyCysPheSerGlnPhe...PheGlyYST 1387
3893 TCGACATGGCATCAGCGCGCGCGGCTTTAGTAGCGGACGCTTCA 3942
|||:::|||||:::|||||
1387 hrGlu.....Ser 1389
3943 GACGGCATCAGAGCAAAATCCGCCCGCGCTGCTGATTCAGCGATTC 3992
|||||:::|||||:::|||||
1390 GlnSerTyrLysAlaLys.....AsnAspValIly 1399
3993 GCGAAGATACCGCGCAGGTTCGCGGATTCGCGATGAAACCGCACATCG 4042
:::|||||:::|||||
1399 sSerTyrMetGlyAlaIleTyrAlaGlyIleLeuAlaGlyPro..... 1413
4043 GCGCAACGCGGTATTCGTCCAAAAAGCGATTTCGATCGCAAAAGCTC 4092
|||||:::|||||:::|||||
1414 .....TrpLeuIleLysGlyAlaPheValTyrGlyAsnIle 1425
4093 AAT.....ATCCGCAACCGCGG 4109
|||||:::|||||:::|||||
1426 AsnAsnAspLeuThrThrAspTyrGlyThrLeuGlyIleSerThrGlyse 1442
4110 CCTTGCAATTACCGCTACCGCGCGGCGCATTAAGCGGATTAATTCATTC 4158
:::|||||:::|||||
1442 rTrpIleGlyLysGlyPheIleAlaGlyThrSerIleAspTyrArgTyrI 1459
4159 .....AAACGCGCGCAACACATTTCC.....ATCAGCGCT 4188
|||||:::|||||:::|||||
1459 LeValAsnProArgArgPheIleSerAlaIleValIleSerThrValAlaPro 1475
4189 TATTGAGCGCTGTCTTATCCGAT.....GCCGCTTC 4220
|||||:::|||||:::|||||
1476 PheValIleIleAlaGlyTyrValArgIleAspLeuProGluIleSerGln 1492
4221 CGGCAAA...GTCCGACAGCGCGCTCAATACCGCGGATTCGCCGAGATT 4267
|||||:::|||||:::|||||
1492 nGlyLysGluValArgThr.....P 1499
4268 TCGGCAAAACCGCGAGTGCAGATGGGGCTAAACCGCAATCAAAAGT 4317
|||:::|||||:::|||||
1499 heGlnLysThrArgPheGluAsnValAlaIle.....Pro 1510
4318 TTCACGCTGTTCCTCCACGCTGCCCGCGCAAGGCGCGCAATTGGAAC 4367
|||||:::|||||:::|||||
1511 PheGlyPheAlaLeuGlnHisAlaTyrSerArgGlySerArgAlaGluVa 1527
4368 GCAGCACAGCGCGCATCAATTAAGCTAC 4398
:::|||||:::|||||
1527 LAsn.....SerValGlnLeuAlaTyr 1534

```

seqname: SwissProt_40:OMP_RICRI

seq_documentation_block:

ID OMP_RICRI STANDARD; PRT; 1654 AA.

AC 053047;

DT 30-MAY-2000 (Rel. 39, Created)

DT 30-MAY-2000 (Rel. 39, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Outer membrane protein B precursor (168 kDa surface-layer protein)

DE (Surface protein antigen) (Cell surface antigen 5) (Scas5) (rompB)

DE (romp B) [contains: 120 kDa surface-exposed protein (surface protein

DE antigen) (120 kDa outer membrane protein ompB); 32 kDa beta peptide].

GN OMPB

OS Rickettsia rickettsii.

OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;

OC Rickettsiaceae; Rickettsiinae; Rickettsia.

OX NCBI_Taxid=783;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-R.

RX MEDLINE-92167802; PubMed-1724278;

RA Gilmore R.D. Jr., Cleplak W. Jr., Policastro P.F., Hackstadt T.;

RT "The 120 kilodalton outer membrane protein (romp B) of Rickettsia

RT rickettsii is encoded by an unusually long open reading frame:"

RT evidence for protein processing from a large precursor.";

RL Mol. Microbiol. 5:2361-2370(1991).

RN [2]

RP SEQUENCE OF 279-1654 FROM N.A.

RC STRAIN-R.

RX MEDLINE-90136087; PubMed-2515418;

RA Gilmore R.D. Jr., Joste N., McDonald G.A.;

RT "Cloning, expression and sequence analysis of the gene encoding the

RT 120 kD surface-exposed protein of Rickettsia rickettsii.";

RL Mol. Microbiol. 3:1579-1586(1989).

CC -1- FUNCTION: THE 120 KDA SURFACE-EXPOSED PROTEIN IS A MAJOR

CC VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION.

CC -1- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A S-

CC LAYER WITH HEXAGONAL SYMMETRY.

CC -1- SIMILARITY: BELONGS TO THE RICKETTSIAE OMPA/OMP FAMILY.

CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation-

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See http://www.isb-sib.ch/announce/

CC or send an email to license@isb-sib.ch).

CC -----

CC DR EMBL; X16353; CAA34403.1; .

DR InterPro; IPR003858; rompa_rompB.

DR Pfam; PF02708; rompa_rompB; 1.

KW Antigen; S-layer; Cell wall.

FT CHAIN 1 1333 120 KDA SURFACE-EXPOSED PROTEIN.

FT DOMAIN 1334 1654 32 KDA BETA PEPTIDE.

FT CHAIN 1181 1188 POLY-THR.

FT DOMAIN 1654 1654 AA; 168184 MW; D7AB70FB7087F618 CRC64;

SQ SEQUENCE

SQ

SQ

SQ

SQ

SQ

SQ

SQ

SQ

SQ

SQ

SQ

SQ

SQ

723 CACAGTCACTTA.....GTAGCGGAAAAAATTT..... 750

724 CACAGTCACTTA.....GTAGCGGAAAAAATTT..... 750

725 CACAGTCACTTA.....GTAGCGGAAAAAATTT..... 750

726 CACAGTCACTTA.....GTAGCGGAAAAAATTT..... 750

727 CACAGTCACTTA.....GTAGCGGAAAAAATTT..... 750

728 CACAGTCACTTA.....GTAGCGGAAAAAATTT..... 750

729 CACAGTCACTTA.....GTAGCGGAAAAAATTT..... 750

730 CACAGTCACTTA.....GTAGCGGAAAAAATTT..... 750

731 CACAGTCACTTA.....GTAGCGGAAAAAATTT..... 750

732 CACAGTCACTTA.....GTAGCGGAAAAAATTT..... 750

733 CACAGTCACTTA.....GTAGCGGAAAAAATTT..... 750

734 CACAGTCACTTA.....GTAGCGGAAAAAATTT..... 750

735 CACAGTCACTTA.....GTAGCGGAAAAAATTT..... 750

736 CACAGTCACTTA.....GTAGCGGAAAAAATTT..... 750

737 CACAGTCACTTA.....GTAGCGGAAAAAATTT..... 750

738 CACAGTCACTTA.....GTAGCGGAAAAAATTT..... 750

739 CACAGTCACTTA.....GTAGCGGAAAAAATTT..... 750

740 CACAGTCACTTA.....GTAGCGGAAAAAATTT..... 750

741 CACAGTCACTTA.....GTAGCGGAAAAAATTT..... 750

742 CACAGTCACTTA.....GTAGCGGAAAAAATTT..... 750

743 CACAGTCACTTA.....GTAGCGGAAAAAATTT..... 750

744 CACAGTCACTTA.....GTAGCGGAAAAAATTT..... 750

745 CACAGTCACTTA.....GTAGCGGAAAAAATTT..... 750

746 CACAGTCACTTA.....GTAGCGGAAAAAATTT..... 750

747 CACAGTCACTTA.....GTAGCGGAAAAAATTT..... 750

748 CACAGTCACTTA.....GTAGCGGAAAAAATTT..... 750

749 CACAGTCACTTA.....GTAGCGGAAAAAATTT..... 750

750 CACAGTCACTTA.....GTAGCGGAAAAAATTT..... 750

751 CACAGTCACTTA.....GTAGCGGAAAAAATTT..... 750

752 CACAGTCACTTA.....GTAGCGGAAAAAATTT..... 750

753 CACAGTCACTTA.....GTAGCGGAAAAAATTT..... 750

754 CACAGTCACTTA.....GTAGCGGAAAAAATTT..... 750

755 CACAGTCACTTA.....GTAGCGGAAAAAATTT..... 750

756 CACAGTCACTTA.....GTAGCGGAAAAAATTT..... 750

757 CACAGTCACTTA.....GTAGCGGAAAAAATTT..... 750

758 CACAGTCACTTA.....GTAGCGGAAAAAATTT..... 750

759 CACAGTCACTTA.....GTAGCGGAAAAAATTT..... 750

760 CACAGTCACTTA.....GTAGCGGAAAAAATTT..... 750

761 CACAGTCACTTA.....GTAGCGGAAAAAATTT..... 750

762 CACAGTCACTTA.....GTAGCGGAAAAAATTT..... 750

763 CACAGTCACTTA.....GTAGCGGAAAAAATTT..... 750

764 CACAGTCACTTA.....GTAGCGGAAAAAATTT..... 750

765 CACAGTCACTTA.....GTAGCGGAAAAAATTT..... 750

766 CACAGTCACTTA.....GTAGCGGAAAAAATTT..... 750

767 CACAGTCACTTA.....GTAGCGGAAAAAATTT..... 750

768 CACAGTCACTTA.....GTAGCGGAAAAAATTT..... 750

769 CACAGTCACTTA.....GTAGCGGAAAAAATTT..... 750

770 CACAGTCACTTA.....GTAGCGGAAAAAATTT..... 750

771 CACAGTCACTTA.....GTAGCGGAAAAAATTT..... 750

772 CACAGTCACTTA.....GTAGCGGAAAAAATTT..... 750

773 CACAGTCACTTA.....GTAGCGGAAAAAATTT..... 750

774 CACAGTCACTTA.....GTAGCGGAAAAAATTT..... 750

775 CACAGTCACTTA.....GTAGCGGAAAAAATTT..... 750

776 CACAGTCACTTA.....GTAGCGGAAAAAATTT..... 750

777 CACAGTCACTTA.....GTAGCGGAAAAAATTT..... 750

778 CACAGTCACTTA.....GTAGCGGAAAAAATTT..... 750

779 CACAGTCACTTA.....GTAGCGGAAAAAATTT..... 750

780 CACAGTCACTTA.....GTAGCGGAAAAAATTT..... 750

781 CACAGTCACTTA.....GTAGCGGAAAAAATTT..... 750

782 CACAGTCACTTA.....GTAGCGGAAAAAATTT..... 750

783 CACAGTCACTTA.....GTAGCGGAAAAAATTT..... 750

784 CACAGTCACTTA.....GTAGCGGAAAAAATTT..... 750

785 CACAGTCACTTA.....GTAGCGGAAAAAATTT..... 750

786 CACAGTCACTTA.....GTAGCGGAAAAAATTT..... 750

787 CACAGTCACTTA.....GTAGCGGAAAAAATTT..... 750

788 CACAGTCACTTA.....GTAGCGGAAAAAATTT..... 750

789 CACAGTCACTTA.....GTAGCGGAAAAAATTT..... 750

790 CACAGTCACTTA.....GTAGCGGAAAAAATTT..... 750

1571 ValTYr...ProGluValHisAla 1577

seq_name: SwissProt_40: FHAB_BORPE

seq_documentation_block:

ID FHAB_BORPE STANDARD; PRT; 3591 AA.

AC P12255;

DT 01-OCT-1989 (Rel. 12, Created)

DT 01-FEB-1996 (Rel. 33, Last sequence update)

DT 01-FEB-1996 (Rel. 33, Last annotation update)

DE Filamentous haemagglutinin.

GN FHAB.

OS Bordetella pertussis.

OC Bacteria; Proteobacteria; beta subdivision; Alcaligenaceae;

OC Bordetella.

OX NCBI_Taxid:520;

RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.

RX MEDLINE-90355839; PubMed-2388559;

RA Reiman D.A., Domenighini M., Tuomanen E., Rappuoli R., Falkow S.,

"Genetic characterization of Bordetella pertussis filamentous

haemagglutinin: a protein processed from an unusually large

precursor.";

RL Mol. Microbiol. 4:787-800(1990).

RM [2]

SEQUENCE OF 1-3261 FROM N.A.

RX MEDLINE-89202384; PubMed-2539596;

RA Reiman D.A., Domenighini M., Tuomanen E., Rappuoli R., Falkow S.,

"Filamentous haemagglutinin of Bordetella pertussis: nucleotide

sequence and crucial role in adherence.";

RL Proc. Natl. Acad. Sci. U.S.A. 86:2637-2641(1989).

-1- FUNCTION: EVIDENCE FOR A ROLE IN HOST-CELL BINDING AND

INFECTION.

-1- SUBCELLULAR LOCATION: SURFACE.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL; M60351; AAA22974.1; -

DR EMBL; M60351; AAA22975.1; ALT_INT.

DR EMBL; M60351; AAA22976.1; ALT_INT.

KW Antigen; Hemagglutinin.

SQ SEQUENCE 3591 AA: 367420 MW; EF7418B30D6E5138 CRC64;

alignment_scores:

Quality: 217.00 Length: 1290;

Ratio: 0.391 Gaps: 62

Percent Similarity: 43.023 Percent Identity: 20.078

alignment_block:

US-09-303-518D-653 x FHAB_BORPE ..

Align seg 1/1 to: FHAB_BORPE from: 1 to: 3591

```

1213 TTGATCTTACAGCAGCATCACCAGCGCGGCGTGTGATTTTGA 1262
    ||| ||||| ||| ||||| |||
2482 LEUTYRLEUTHIRGLUALATHIRAGLNLILYLEHTRASPGLNTRYALAGI 2498
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1263 GGGTAATTTTTCAGGTCTGCGCTAAACACGAAACGTCGCAAGCGCGG 1312
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2498 YGLYGLYALALEUILLEA.....SerglycylA 2508
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1313 GCGTTCATATC...AGTGTGCGAGTACCGTCTACTTGAAAGTAAGCCG 1359
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2508 SPVALTHVALASNTHRASPELTHASPAVALSER...SerValAsnGly 2523

```

```

1360 GTGGCAACGACCGCTGTCCAAATC.....GCCAAAGCAGCGTCT 1403
    : : : : : ||| : : : : : ||||| : : : : :
2524 LEULIEGLIYARGSERVALYSVALASPALAGLYSGLYLVALVA 2540
    ||| : : : : : ||||| : : : : : ||||| : : : : :
1404 GGT...CAAGCCAAAGGGGAAACCAAGCGTGGTCAGCGGCGGACG 1450
    ||| : : : : : ||||| : : : : : ||||| : : : : :
2540 IVALAASPSERLYSGLY.....AlaGly 2549
    ||| : : : : : ||||| : : : : : ||||| : : : : :
1451 GTRAAAGTCATCTTAATCAGCAGCGGAGCATCAAGCCAAACACGCC 1500
    || : : : : : ||||| : : : : : ||||| : : : : :
2549 IYGLIYIEGLIUALASPSRGLVALASPAVALSERGLYARG..... 2562
    || : : : : : ||||| : : : : : ||||| : : : : :
1501 TTTAGTAATGCGCTGTGTCAGCGC.....AGGGG...ACGGTCA 1541
    : : : : : ||||| : : : : : ||||| : : : : :
2563 .....ASPLIEGLIYIEGLIYLYSLEUADYGLIYASPAVALR 2577
    : : : : : ||||| : : : : : ||||| : : : : :
1542 ACTGAATGCCAT.....AATCAGTTCACCCGCCACA 1573
    : : : : : ||||| : : : : : ||||| : : : : :
2577 GLEULYSALASPTHRVALIYSLVALATHSERMETARGTYRSPASPL 2594
    : : : : : ||||| : : : : : ||||| : : : : :
1574 AACTGTATTTGGCGTTTCGCGGC...GGACGTTGGATTGACGGGCAT 1620
    || : : : : : ||||| ||| ||||| : : : |||
2594 YSGLYARGLEUALAALARGGLYASPGIYALALEUASPALAGLNGLY 2610
    : : : : : ||||| : : : : : ||||| : : : : :
1621 TCGCTTGTCCACCGCATTCAAATACCGATGAAGGCGGATGATTGT 1670
    : : : : : ||||| : : : : : ||||| : : : : :
2611 GINLEUHIISLIEGLIUALALYSARGLEULIYTHRALAGLYALA..... 2624
    : : : : : ||||| : : : : : ||||| : : : : :
1671 CAACCAATCAGACCAAGAAATCCACCGTTACCATTCAGCAATTA 1719
    : : : : : ||||| : : : : : ||||| : : : : :
2625 .....Thleuysgllylylv 2631
    : : : : : ||||| : : : : : ||||| : : : : :
1720 .....GATATTACTACACCGGC.....AATPAC 1743
    : : : : : ||||| : : : : : ||||| : : : : :
2631 ALYSLEUASPAVALASPAVALIYLSLEUGLYLYVALIYGLIUALAGLY 2647
    : : : : : ||||| : : : : : ||||| : : : : :
1744 AACCACTGTGATAGCAAAAGAAATGGCTACACGGTGTGGC... 1791
    : : : : : ||||| : : : : : ||||| : : : : :
2648 SERSETYRGLIUDASNLYSERSETHRPROLEUYSERLEUPHEALAI 2664
    : : : : : ||||| : : : : : ||||| : : : : :
1792 GAGAAATGTCACCAAAACGACGCGGCTCAATTCATTAATCAAC 1840
    : : : : : ||||| : : : : : ||||| : : : : :
2664 ELEUSERSETHRTHRGITHRASNGINSERALAHISALASNHISLYRG 2681
    : : : : : ||||| : : : : : ||||| : : : : :
1841 CGGAAGAACCGCAT.....CCCACTTACTG 1866
    : : : : : ||||| : : : : : ||||| : : : : :
2681 LYTHIRARGIIEGLIAGLYTHIRLEUGLIYLYSMETGLIASNLEUGLY 2697
    : : : : : ||||| : : : : : ||||| : : : : :
1867 CTTTCGCGCGGACACA.....AATTTAAGCGCAATATCACGCAACAA 1910
    : : : : : ||||| : : : : : ||||| : : : : :
2698 IIEGLIYGLISERVALASPAVALAHISHTHRASPLEUSERVALAIAT 2714
    : : : : : ||||| : : : : : ||||| : : : : :
1911 CGGCAAACTGTTTTCAGCGGACAGCCGACGCGCCTTCAAT... 1956
    : : : : : ||||| : : : : : ||||| : : : : :
2714 GASPAALARGPHELYSALALALASRPHEDALAHISALAGLIHISGLUL 2731
    : : : : : ||||| : : : : : ||||| : : : : :
1956 ..... 1956
2731 YSASPAVALARGINLEUSERLEUGLYALALYSVALGLYALGLYLYR 2747
    : : : : : ||||| : : : : : ||||| : : : : :
1957 .....CATTTAGG 1964
    : : : : : ||||| : : : : : ||||| : : : : :
2748 GLUALAGLYPHESEULEUGLYSERGLIUSERGLYLEUGLUALAHISLAGI 2764
    : : : : : ||||| : : : : : ||||| : : : : :
1965 AAGCGGTGTGTCAAAATGAAAGGTATCCACAGAGAAATCGTG... 2010
    : : : : : ||||| : : : : : ||||| : : : : :
2764 YARGGLYMETHTHRLAGLYALAGIUALIYSLVALGLYTYRAGDALASERN 2781
    : : : : : ||||| : : : : : ||||| : : : : :
2011 ..TGGAACAAGATGATGACGACCGACATTTAAACGGAAACCTTCAT 2058
    : : : : : ||||| : : : : : ||||| : : : : :
2781 ISGLIUGINSESRGLIUTHIRGLIYLSERSETYRARGASNLASNLLEUASN 2797
    : : : : : ||||| : : : : : ||||| : : : : :
2059 ATTACAGGCGGACAA.....GC 2075

```

```

2798 PheGlyGlySerValGluAlaGlyAsnValLeuAspIleGlyAla 2814
2796 GGTGGTTCCCGCAAT..... 2091
2814 AspIleAsnIleArgAsnArgTyrGlyGlyAlaIleAlaGlySerAlaGlyTyr 2831
2092 ..... GTTCCCAAGTGGAGCGCATTTGGCAT 2118
2831 hrGluAlaIleuArgMetArgMetAlaLysLysValGluSerThrLysTyr 2847
2119 TTAAAGCATACGCCCAAGCAGTTTTCGGTGTGCGACCGCATCAAGCA 2168
2848 ValSerGluGlnThrSerGlnSerSerGlyTyrSerValGluValAlaIle 2864
2169 CACAATCTGTAACAGTTTCGGACTGGACGGGTCTGACAAAGTTTACCGAAA 2218
2864 rThrAlaSerAlaArgSer.....SerLeuLeuThrAla 2876
2219 AAACCATTAACGAGATTAAGTATGCTTCATTGAGCAAG..ACCGAC 2265
2876 IatThrAlaGluGlyAspSerValAlaGlnAsnValGluAspGlyArgGlu 2892
2266 ATCAGAGGCAATGTCAGCCTTCCGATCAGCGCTCATTTAAATCTCAGAG 2315
2893 IleArgGlyGluLeuMetAlaAlaGlnValAla..AlaGluAlaThrGlu 2908
2316 ACTTGCCACA.....CTCAGCGCAATCTTAGTG 2344
2908 nLeuValThrAlaAspThrAlaAlaValAlaLeuSerAlaGlyIleSerA 2925
2345 CAGCGGAGACAGCATATACGGTTACGGCAACGCCACCGCAACAAAC... 2391
2925 IaAspPheAspSerSerHis.....SerArgSerThrSerGlnAsnThr 2939
2392 .....GGCACCTCAGC..... 2403
2940 GlnTyrLeuGlyGlyAsnLeuSerIleGluAlaThrGluGlyAspAlaThr 2956
2404 .CTCGTGGGC..... 2412
2956 rLeuValGlyAlaLysPheGlyGlyLysPheGlnValSerLeuLysAla 2973
2413 .....AATGCCAAGCAACATTTAATCAAGCCACA 2442
2973 IaLysSerValAsnLeuMetAlaAlaGluSerThrPheGluSerTyrSer 2989
2443 TTAAAGGCAACACATCGGCTTCGGACAATGCTTCAATTAATCTAAGCA 2492
2990 GluSerHisAsnPheHisAlaSerAlaAspAla.....AsnLeuGlyAla 3004
2493 CAACGCCGACAAACAGCGAGTCAGCGTTTCCGACCAACGCTAAGGCA 2542
3004 AsnAlaValAlaGln..GlyAlaValAlaGlyLeuGlyLeuThrAlaGlyMetG 3020
2543 ACCTAAGCATTCGCCACTCAAC.....GGC 2568
3020 LyThrSerHisGlnIleThrAsnGluThrGlyLysThrTyrAlaGlyThr 3036
2569 AATGTCCTCCCTAGCCGATAGGCAATTCATTTTGAAGACGGCGGT 2618
3037 SerValAspAlaAlaAsnValSerIleAspAlaGlyLysAspLeuAsnIle 3053
2619 TACCGGAAATACAGCGCGGCAAGATACGCATTACACTTAAGACA 2668
3053 uSerGlySerArgValArgGlyLysHisValAlaLeuAspValGlu... 3068
2669 GCGAATGACGCTGCCGTCGGCACGGAATTAAGCAATTTAACTTGAC 2718
3069 .....GlyAspIle..... 3071
2719 AACGCCACCATTAACACTCAATTCGCCCTATCCAGACGATGCGGACGGCG 2768
|||||

```

```

3072 AsnAlaThrSerLysGlnAspGluArgAsnTyrIleAsnSerSerGlyGly 3088
2769 GCAAAACCGCAATGCGGAGATGCGCGCGCGCGCTTCGGCGCGCTTCC 2818
3088 yThrAspAlaSerAlaGlyValAlaIleGlnAsnArgThr..... 3101
2819 rATTAACGCTTACGCCCAACTTCGCGCAAGATCCCTTTCAACAGCTG 2868
3102 .....LeuValAlaProValGlySerAlaGlyPheAsnPheAsnThrGlu 3116
2869 ACGGTAACCGCAATTAAGCATGAGGACATTCCTTTATGTCGGA 2918
3117 HisAspAsnSerArgLeuThrAsnAspGlyAlaAlaGlyValAlaIle 3133
2919 A.....CTCTTCGGCTACCGCGGCAATTTGAAGCTGGCGGAAGTT 2962
3133 rAspGlyLeuThrGlnHisValLysGlyAspAlaAsnLeuThrGlyAlaT 3150
2963 CCGAAGCACTTACACCTTGCTGTCACAAATACCGGCAACGACCGGTA 3012
3150 hr..... 3150
3013 AGTCTCGACCAATTAAGCGTAGTGAAGAAAGACAACACACCGCTGTC 3062
3151 .....IleAlaAspLeuSerGlyLysGlyAsn..... 3159
3063 CGAAATCTTAATTTCACCTTCGCAAAACGACACGTCGATGCCGCGCAT 3112
3159 ..... 3159
3113 GCGGTTATCAGCTTATCCGCAAGACGGGAGATTCGCGCTGATATCCG 3162
3160 .....LeuLysValAspGlyAlaValAlaAsnAlaGlnAsn... 3170
3163 GTCAAAGAACAGAGCTTCCGACAACTCGCGCAAGCGGAGAAACAGA 3212
3171 .....LeuLysAspTyrArgAspLysAspGlyLysSerGlyGlyLeuAs 3185
3213 GCGCGCCTTGACG.....GCAAAACAGCGCAACCTTCCGCCCA 3250
3185 nValGlyIleSerSerThrThrLeuAlaProThrValGlyAlaIlePheG 3202
3251 AACACAGCGGGAAGAAACACACGCGCAAGCCTTACCGCGCTGATTCG 3300
3202 LyArgValAlaGlyGluAsp.....TyrGln 3210
3301 GCGGCGCGCAATGCCAACGAAAGGCAAGATGTTCCGAACGGCGCG 3350
3211 AlaGluGlnArgAlaThrIleAspValGlyGlnThrLysAspProAlaT 3227
3351 G...CAGCGACGGCGGA..... 3365
3227 gLeuGlnValAlaGlyGlyValLysGlyThrLeuAsnGlnAspAlaIleG 3244
3366 .....AATGCCGCAATTAATGACGCGGAGGAGAGAGA 3398
3244 lAlaIleThrValAlaGlnArgAsnLysHisThrAlaGlyGlySerGlu 3260
3399 AAAACGGGTGACGGCGGATTAAGACACACCGCTTGGCAACA.....GC 3442
3261 PheSerValAlaGlyLysSerLeuLysLysAsnGlnValAlaArgProVa 3277
3443 GCGAAGCGGAACCGCGCGGTACACCGCTTCCCGCGCGCGCGCG 3492
3277 lGluThrProThrProAspValAlaAspGlyProProSerArgProThrT 3294
3493 GCGCGCGGATTTGCGCGCAACG.....CA 3518
3294 hrProProAlaSerProGlnProIleArgAlaThrValGluValSerSe 3310
3519 GCCCAACCGCAACCCCAACCGGACGCGCATGATCAGCGCTTAAGCCA 3568
3310 rProProProValSerValAlaThrValGluValAlaProArgProLys 3326

```



```

3569 ATAGCGTTGAGTGAATTTCCGCCACGCTCAACAGCGTTTCGCCGTA 3618
      ::::: ||||| :::::
3327 ValGluThrIaGlnProLeuProIaGlnProValAla..... 3339
3619 CAGAGCAATTTGACCGCGTGTTCGAGACCGCGCCGACGCGCTTG 3668
      ::::: |||||
3340 .....AlaGlnValValProValThr..... 3346
3669 GACAGCGCATCCGGACACCAACACTACCTTCGCAAGATT..... 3712
      ||| ::::: |||
3347 .....ProIaGlnValGlnValAlaValAlaGlnVal 3357
3713 .....TCGCGCGCTACCGCCACCAACAGCGAG 3738
3358 ValProIaGlnProValGlnThrIaGlnProLeuProIaGlnProVal 3374
3739 CTCGCGCAATTCGCTATGCAAGAAAACCTCGACGCGCGCTCGGCAT 3788
      ::::: |||
3374 ValAlaGlnVal.....ThrThrProAlaValGlnProGlnLeuA 3389
3789 CCTGTTTCGACACCGACCGCAACACTTCGACGACG..... 3829
      ::::: |||
3389 IalysValGlnThrValGlnProValLysProGlnThrLysProLeu 3405
3830 .....GCATCGCGCACTCGACGCGCTTCGCCACGCT 3861
3406 ProLysProLeuProValAlaLysValThrLysAlaProProValAla 3422
3862 G.....CCGTTTCGCGCACTACCGCGCGCGAG 3890
3422 IgluThrIaGlnProLeuProValLysProGlnLysAlaThrProG 3439
3891 GTTCGACATCGGCATCGCGCGGCTTTAGTAGCGGACGCTTT 3940
      ||| ::::: |||
3439 LysProValAlaGlnValGlnLysAlaThrVal..... 3449
3941 CAGAGCGCATCGACGCAAAATCCCGCGCGCTCGCATTCAGCGCAT 3990
      ||| ::::: |||
3450 ThrThrValGlnValGlnSerAlaProProLysProAla..... 3462
3991 CAGGCAAGTACCGCGAGTTTCGCGGATTCGCGCATCGCAACCGCAT 4040
3462 ..... 3462
4041 CGGCCCAACGCGCTATTTCCTCAAAAACGAGATTACCGATACGAAAACG 4090
      ||| ::::: |||
3463 .....ProValAlaLysGln 3467
4091 TCATATATCGCACCCCGCGCTTCGATTCACCGCTACCGCGCGCAT 4140
      ||| ::::: |||
3467 nProIaGlnProAlaProLysProLysProLysProLysAlaGln 3483
4141 AAGGAGATTTATTCATCAACCGCGCGCAACATTCATTCACGCGCTTA 4190
      ::::: |||||
3484 ArgPro.....LysProGlnLys..... 3489
4191 TTGGAGCTTCCTATACCGATCCCGCTTCGCGCAACGCGCGCG 4240
      |||
3490 .....ThrThrProLeuSerGlnValValGlnG 3501
4241 TCATATACCGCGCTATTG.....GGCAGGATTCGCGCAACCGCGCG 4281
      ::::: |||||
3501 InGlnValGlnValLeuGlnValGlnAlaSerAlaLysThrLys 3517
4282 AGTGGCGAATGGGC..... 4296
3518 SerLeuProGlnLysLeuProLysProValThrValLysLeuThrAs 3534
4297 .GTAAGCGCGAAATCAAGGTTACGCTTCCTCCACGCTCCCGCG 4345
      ||| ::::: |||
3534 pGlnAsnGlnLysProGlnThrThrThrIleAsnArgArgGlnAsnLeu 3551

```

```

seq_name: SwissProt_40:HX2_HAELN
seq_documentation_block:
ID HX2_HAELN STANDARD; PRT; 928 AA.
AC P45354;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Heme/hemopexin-binding protein precursor (Heme:hemopexin utilization
  protein A).
GN HXUA.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 22-35.
RC STRAIN-DL42 / SEROTYPE B.
RX MEDLINE=95115556; PubMed=7815944;
RA Cope L.D., Thomas S.E., Latimer J.L., Slaughter C.A.,
  Mueller-Eberhard U., Hansen E.J.;
  "The 100 kDa haem:haemopexin-binding protein of Haemophilus
  influenzae: structure and localization."
  J. Microbiol. 13:863-873(1994).
RL [2]
RN SEQUENCE OF 1-30 FROM N.A.
RP STRAIN-DL42 / SEROTYPE B.
RX MEDLINE=95270579; PubMed=7751272;
RA Cope L.D., Yogev R., Mueller-Eberhard U., Hansen E.J.;
  "A gene cluster involved in the utilization of both free heme and
  heme:hemopexin by Haemophilus influenzae type b."
  J. Bacteriol. 177:2644-2653(1995).
RL [1]
CC -1- FUNCTION: BINDS HEME/HEMOPEXIN COMPLEXES.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
CC DR EMBL: U08348; AAA74138.1; -
CC KW Transport; Signal; Repeat.
CC FT SIGNAL 1..21
CC FT CHAIN 22..928 HEME/HEMOPEXIN-BINDING PROTEIN.
CC FT DOMAIN 101..679 6 x 6 AA APPROXIMATE REPEATS.
CC FT REPEAT 101..106 1-1.
CC FT REPEAT 205..210 1-2.
CC FT REPEAT 279..284 1-3.
CC FT REPEAT 410..415 1-4.
CC FT REPEAT 635..640 1-5.
CC FT REPEAT 679..679 1-6.
CC FT DOMAIN 149..172 4 x 6 AA APPROXIMATE TANDEN REPEATS.
CC FT REPEAT 149..154 2-1.
CC FT REPEAT 155..160 2-2.
CC FT REPEAT 161..166 2-3.
CC FT REPEAT 167..172 2-4.
CC FT SEQUENCE 928 AA; 101228 MW; 67D45466A4B92390 CRC64;

```

alignment_scores: Quality: 215.50 Length: 1142
Ratio: 0.457 Gaps: 57

Percent Similarity: 41.331 Percent Identity: 19.527

alignment_block:

US-09-303-518D-653 x HXA2_HAEIN ..

Align seg 1/1 to: HXA2_HAEIN from: 1 to: 928

```

451 AAGCGCATCTTATGGCGCGATTCATATCCCGCGTTGCACAAATT 500
    ::::| | | | | | | | | | | | | | | | | | | | | |
16 SerGlyAlaAlaIleThrProAspPheProGlnHisIleIle 32
    ::::| | | | | | | | | | | | | | | | | | | | | |
501 TGTACAGATGACAGAACCTGTGATGACAGTATATGATGGGTGCA 550
    ::::| | | | | | | | | | | | | | | | | | | | | |
32 rValPheGlyThrValThrIleGluIleIleIleIleIle 43
    ::::| | | | | | | | | | | | | | | | | | | | | |
551 AATACGCTGAT.....TTAAATAAATACCTGATCGTGTGCATAC 591
    ::::| | | | | | | | | | | | | | | | | | | | | |
44 ....AlaAspLysMetThrIleIleGlnIleSerAspLysAlaGlnIle 58
    ::::| | | | | | | | | | | | | | | | | | | | | |
592 GGAGCGAGCAGACATTTGGCGGTCTGATGAA..... 624
    ::::| | | | | | | | | | | | | | | | | | | | | |
59 Asp.....TrrLysSerPheAspLleGlyGlnLysGln 70
    ::::| | | | | | | | | | | | | | | | | | | | | |
625 .....GACGACCAATACCGCGAAAGTTTCATATCATATTGCAA 664
    ::::| | | | | | | | | | | | | | | | | | | | | |
70 uValLysPheGlnIleProAsnGlnHis.....AlaVal 81
    ::::| | | | | | | | | | | | | | | | | | | | | |
665 GCGCATATTTCTGGCTGCTGCGTGCGCAATACCTTGCACAAATGCA... 711
    ::::| | | | | | | | | | | | | | | | | | | | | |
81 alaIaIleAsnArgValIleGlyGlnLysAlaSerGlnIleGlnIleLys 97
    ::::| | | | | | | | | | | | | | | | | | | | | |
712 ...TCAGTGGTGGCAGACGTCACACTAGTAGGCAAAATTAACATAG 758
    ::::| | | | | | | | | | | | | | | | | | | | | |
98 LeuThrAlaAsnGlyLysValTyrLeuAla.....As 108
    ::::| | | | | | | | | | | | | | | | | | | | | |
759 CCCATATGATTTTTCACCAACAGAGGCTCATTTGGCAGACGTGGCTCAC 808
    ::::| | | | | | | | | | | | | | | | | | | | | |
108 nProAsnGlyValIleIleThrGlnIleAla..... 118
    ::::| | | | | | | | | | | | | | | | | | | | | |
809 CAATGTTATCTATGATGCGCAAAAGCAAAAGTGTTAATTAATGGGCTA 858
    ::::| | | | | | | | | | | | | | | | | | | | | |
119 .....GlnIleAsnValAlaIleGlyLeu 125
    ::::| | | | | | | | | | | | | | | | | | | | | |
859 TTGCAACAGCAGCAACCCCTATATAGAAAGCAATGGCTCCAGTACGTAGT 908
    ::::| | | | | | | | | | | | | | | | | | | | | |
126 LeuAlaThrThrLys.....AspLeuGln 133
    ::::| | | | | | | | | | | | | | | | | | | | | |
909 TCGTAAAGATTGGTTCTATGATGAATCTTGTGAGATACCCATTACAG 958
    ::::| | | | | | | | | | | | | | | | | | | | | |
133 uArg..... 134
    ::::| | | | | | | | | | | | | | | | | | | | | |
959 TATTCACGACACATCAAAATGGAATACTTTTAAACGACATTAAT 1008
    ::::| | | | | | | | | | | | | | | | | | | | | |
135 .....IleSerGlnAsnSer 139
    ::::| | | | | | | | | | | | | | | | | | | | | |
1009 AATGCGCAGAGAAAATGATGCGCAAAACATAACATAATCTCTACCTTA 1058
    ::::| | | | | | | | | | | | | | | | | | | | | |
140 AsnSer.....TyrGlnPheThr 146
    ::::| | | | | | | | | | | | | | | | | | | | | |
1059 TAGATTAAACACGACCGTTCAATTTTAATGTTTCTTTATCCGAGA 1108
    ::::| | | | | | | | | | | | | | | | | | | | | |
146 gArgThrLysAspArgGlnVal..... 153
    ::::| | | | | | | | | | | | | | | | | | | | | |
1109 CAGCAAGAGAACCTGTATCATGCTGCAGGTGGGTCAACAGTTATGCA 1158
    ::::| | | | | | | | | | | | | | | | | | | | | |
154 ..LeuLysGlnGlyLeuValLeuLysAspGlyGlnValValLysGlnGly 169
    ::::| | | | | | | | | | | | | | | | | | | | | |
1159 CCCAGACTGAATATGAGAAAATATTTCC.....TTTATGACAA 1199
    ::::| | | | | | | | | | | | | | | | | | | | | |
170 GlnValIleAsnGlnGly...AsnIleThrAlaGlnAspPheValIle 185
    ::::| | | | | | | | | | | | | | | | | | | | | |
1200 AGGAAAGAGTGAATGATCTTACAGACATCAACAGGCGGGGCG 1249
    ::::| | | | | | | | | | | | | | | | | | | | | |

```

```

185 uAsnGlyAspClnValIleAsnLysGlyAsnIleAsn..... 197
    ::::| | | | | | | | | | | | | | | | | | | | | |
1250 GTTGTATTTTGGAGGCTAATTTTACGGCTCGCCCTAAAAACACGAAAGC 1299
    ::::| | | | | | | | | | | | | | | | | | | | | |
198 .....ValGlnLysAsnSerThrIleAsnGlnLys..... 207
    ::::| | | | | | | | | | | | | | | | | | | | | |
1300 TGGCAAGCGCGCGCTTCATATCATGATGACGATGACCTTACTTGAA 1349
    ::::| | | | | | | | | | | | | | | | | | | | | |
208 .....ValTyrLeuSerSerGlyIleAsnPheThr 219
    ::::| | | | | | | | | | | | | | | | | | | | | |
1350 AGTA.....AAGCGCTGCGCAACGACCGCTGTCCAAATTCGGCAAG 1393
    ::::| | | | | | | | | | | | | | | | | | | | | |
219 rLeuProAspSerGlyIleSer.....ValAlaLeuG 230
    ::::| | | | | | | | | | | | | | | | | | | | | |
1394 GCACCGCTGCTGTCAAGCAAGAGGCAAAACGACGCTCGTCAAGCTG 1443
    ::::| | | | | | | | | | | | | | | | | | | | | |
230 uAspAsnThrValIleGlnGlyIleValLysAsnGlnGlySerIleLysAla 246
    ::::| | | | | | | | | | | | | | | | | | | | | |
1444 GCGCAGCGTAAAGTCATCTTATGATCAGCAGCGGACGATCAAGCAAAA 1493
    ::::| | | | | | | | | | | | | | | | | | | | | |
247 GlyGln.....IleThrLeuSerAla.....LysGlyArgLys 257
    ::::| | | | | | | | | | | | | | | | | | | | | |
1494 ACAAGCTTATGTAATCGCTTGTGTCAGCGGC..... 1527
    ::::| | | | | | | | | | | | | | | | | | | | | |
257 sGlnAlaLeuAspSerLeuValMetAsnAsnGlyValLeuGlnAlaThrL 274
    ::::| | | | | | | | | | | | | | | | | | | | | |
1528 .....AGGCGAGCGTGCACACTGAATGCGCATTAAT...CAG 1560
    ::::| | | | | | | | | | | | | | | | | | | | | |
274 yValSerAsnLysAsnGlyLysValValLeuSerAlaAspAsnValGln 290
    ::::| | | | | | | | | | | | | | | | | | | | | |
1561 TTCAACCCCGCAAACTATTTGCGCTTTCGCGCGGACGTTTGATTT 1610
    ::::| | | | | | | | | | | | | | | | | | | | | |
291 LeuAsnAsnGlnSer.....AsnIle 297
    ::::| | | | | | | | | | | | | | | | | | | | | |
1611 GAACGGCGCTTGGCTTGGTTCACCGCATTCAAATACCGATGAAGGG 1660
    ::::| | | | | | | | | | | | | | | | | | | | | |
297 eLysGlnGlnIleValThrPhe.....GlyA 306
    ::::| | | | | | | | | | | | | | | | | | | | | |
1661 CGATGATTTGTCAACACATCAAGCAAGAAATCCACCGTTACATTACA 1710
    ::::| | | | | | | | | | | | | | | | | | | | | |
306 lAspValAlaThrSer...AsnLysGlnLeuLysAspAsnIleLysIleThr 321
    ::::| | | | | | | | | | | | | | | | | | | | | |
1711 GGCATTAAGATATTACTACACCGGCAATATACACACTTGATAGCA 1760
    ::::| | | | | | | | | | | | | | | | | | | | | |
322 SerLysThrGlySerLysValThrSerProLysIleAsnPhe...ThrGln 337
    ::::| | | | | | | | | | | | | | | | | | | | | |
1761 AAAAGAAATTCCTACACGCTTGGTTGGCGAGAAAGATGCAAC... 1806
    ::::| | | | | | | | | | | | | | | | | | | | | |
337 yLysSerValAsnIleAsnGlyAsnPheGlyArgGlnAspSerThrThr 354
    ::::| | | | | | | | | | | | | | | | | | | | | |
1807 .....AAACGACGCGCGCTCATCTGAATTAC 1836
    ::::| | | | | | | | | | | | | | | | | | | | | |
354 lAspValAlaThrSerPheLysLysLeuAsnThrGlnValAsnIleAspVal 370
    ::::| | | | | | | | | | | | | | | | | | | | | |
1837 CAACCGGAGAA.....GGGATCGGACT... 1860
    ::::| | | | | | | | | | | | | | | | | | | | | |
371 ProAspAsnGlnAsnIleArgIleAlaAspIleGlnuAspAsnThrGlyThr 387
    ::::| | | | | | | | | | | | | | | | | | | | | |
1860 ..... 1860
    ::::| | | | | | | | | | | | | | | | | | | | | |
387 rGlyThrThrGlyThrGlyThrSerSerPheIleGlnThrGlyAlaLeuS 404
    ::::| | | | | | | | | | | | | | | | | | | | | |
1861 .....TTACTGCTTTCCGCGGAGAACAAATTTAAACGGCAT... 1896
    ::::| | | | | | | | | | | | | | | | | | | | | |
404 eSerLeuLeuAlaAsnAsnGlyLysValAsnLeuLysGlyAsnAsnVal 420
    ::::| | | | | | | | | | | | | | | | | | | | | |
1896 ..... 1896
    ::::| | | | | | | | | | | | | | | | | | | | | |
421 AsnIleSerGlyArgIleHisIleAspSerPheArgGlySerAspSerle 437
    ::::| | | | | | | | | | | | | | | | | | | | | |
1897 .ATACGCAACAAC..... 1911
    ::::| | | | | | | | | | | | | | | | | | | | | |
437 uLeuLysLeuThrAsnLysGlyHisIleAspIleAsnAsnAlaAspIleH 454
    ::::| | | | | | | | | | | | | | | | | | | | | |

```

```

1912 .....GGCAACTGTTTTC..... 1926
1913 .....|||||.....
454 isserlysglyargleuphephelethrserleuclnslngluasp 470
1927 .....AGCGGACGACGACGACGCGCTTCAATCATTTAGGAAGCG 1970
1928 .....|||||.....
471 pheylsersanlethrlethrapsrlyslleasleuclnslngluasp 487
1971 G.....T 1972
487 yalameclglyleucllyargservalasplulysaspyraspasnargt 504
1973 GGTCAAAAATGGAAGGTATCCCAAGAGA.....CAATCGTGGGAGC 2016
1974 .....|||||.....
504 rpolnlystrngllyserglngllysllyspheaspvallysmetser 520
2017 AAC.....GATTGATCGACGCGCATTTAAAGC..... 2046
2018 .....|||||.....
521 asnvalglupheasnnglnvalaspsvalilleuallaglylphedl 537
2047 .....GAAACTCCATA 2059
537 ulysvalasleuaspplysillevalalathrglylnthrasnpheyr 554
2060 TTCAGGCGGACAAAGCGGTGTTCCCGCAATGTGCCAAAGTGGAA... 2106
2061 .....|||||.....
554 leaspllygly.....ValserArgasnnglyarglystrylunyr 567
2107 GCGCATTTGACATTTAAGCAATGACGCCCAAGCA..... 2139
2108 .....|||||.....
568 glyvalleuaspluasplysargthrnglneusergluleuasngl 584
2140 .....GTTTCGGTGCACCGCATCCCAAGCCACA 2170
584 yargargargtrpglytrlytrlyraspluclneuasplmetasna 601
2171 CATCTGTACAGTTCGACGTGACGCGGTGCACAAGTTGACGAAAAA 2220
2172 .....|||||.....
601 latyrlleuyltrpghaspheasplneualthrlyasnthrlyargser 617
2221 ACCATTACCGAC..... 2232
2222 .....|||||.....
618 thrilleysasprthrngluleasnllesersnsersanlleasleu 634
2233 .....GATTAAGTATTCCTCATTTAGACA 2257
2234 .....|||||.....
634 sasnglyphevalhlleuallagluylslleusleuaspsnsert 651
2258 AGACGACATC.....AGAGCAATGTACGCTTGGCGAT 2292
2259 .....|||||.....
651 ysileaspllethrheasplysaspsnserserlnasprthrleual 667
2293 CACGCTCATTTAAATCTCACAGA...CTGCCACATCAACGCGCAATCT 2339
2294 .....|||||.....
668 thrasnargleuclmetasnnglylvalaslermetlleasnsers 684
2340 TAGTCACGCGGACGACGACGCTTACGCTTACGCGCAACGCCAACA 2389
2341 .....|||||.....
684 elysllevalaglyasplulysglnglylleserprothrnglythry 701
2390 ACGCAACCTC.....AGCTCGTGGGCAATGCCCAAGCAACATTT 2430
2391 .....|||||.....
701 latnmepheleuileglyleuclleuglyllysserserlelph 717
2431 AATCAAGC.....ACATTAACGCAACACATCGCGC 2462
2432 .....|||||.....
718 valysserhlsgllytrlytrlypheylstraspplysanthrlysl 734
2463 TTCGACAAATGCTCA.....TTTAAATTA 2488
2464 .....|||||.....
734 ealeglyllystrlyserlysgluasplulysllethrallaleasn 751

```

```

2489 GCAACACCGCGTACAA.....AACGCACTGTGACGCTTTC 2526
2490 .....|||||.....
751 lyglyarglalaalaglulvalleuileasnglylaleucllyserala 767
2527 GACACGCGTAAGCAACGTAAGCATTCGCGCATTCGCAACGCGCATTCG 2576
2528 .....|||||.....
768 aspsan...AspAlaasnillealasnmetalaphehtrilleglyaspse 783
2577 CCGACCGCATTAAGCGCATTCATTTTGAAGACGCGCTTACCGGAA 2626
2578 .....|||||.....
783 ralaasnthrlystrthr.....llegluasnlalasphelethrall 798
2627 AATTCAGCGGCGGCAAGATACGCGCATTTACCTTAAAGACGCGAATG 2676
2628 .....|||||.....
798 euallaproasnnglyllythrlyleucllyserlyaspsvalglu... 813
2677 ACGCTGCGCGTGGGACGCGAATTAGCAATTTAACTTGACACGCGCAC 2726
814 .....llegl 815
2727 CATTAACCTCAATTCGCGCTATCGACACGATCGCGGCGCGCAACCG 2776
2728 .....|||||.....
815 uvallyspromsersasnphe..... 822
2777 GCAGCGCGGACGATGCGCGCGCGCGCGCTTCCCTTATTC 2826
823 .....Thrlypnegluleuproarg..... 829
2827 GTTACGCGCGCAACTGCGCAATTCGCTTCAACACGCTGACGCTAA 2876
830 .....Glulysasnleuasnlthrlyslleas 839
2877 CGGC.....AATTTGAACGCTCAGGACATTCGCTTATGTCGG 2917
839 .....|||||.....
839 ngllyalaserthrlyleucllyserglnglylphelalargleuylraspl 856
2918 AACTCTTCGCGTACCGGACGCGCAATTTGAAGCTTGGCGGAAGTTCCGA 2967
2919 .....|||||.....
856 ysileasnnglylvalarglaserasnleu..... 865
2968 GGCACCTTACACTTGGCTGTCAACATATCCGCGGACGACCGTAGTCT 3017
866 .....Serl 867
3018 CGAGCAATTCAGGCTAGTAGAAGAAAGAC.....AACACACCGC 3058
867 .....|||||.....
867 agluclneuasnlvalthrapsalaserglulyslleleasnthrlysl 884
3059 TGTCCGAATACTTAATTC.....ACCTTCGCAAAACGAACACGTCGAT 3102
3100 .....|||||.....
884 euvalserleuaspvalglulysleuvalservalalavalcysasp 900
3103 GCGCGCGCATGCGCTTATCAGCTTATCCGCAAGACGCGGAGTCCGCT 3152
3104 .....|||||.....
901 Alegly..... 902
3153 GCATTAATCCGCTCAAGAACAGAGCTTTCGCAACATCGGC..... 3195
3154 .....|||||.....
903 .....Asnnglycysglnglclnlnpheglyaspplysglyasnthrly 918
3196 .....AAGCGGGAAGAAACAGAGCC 3216
3197 .....|||||.....
918 ysvalservalglyluleucllyala 926

```

seq_name: swissProt_40:HLXA_SERMA

seq_documentation_block:

ID HLXA_SERMA STANDARD; PRT; 1608 AA.
AC P15320;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 01-NOV-1990 (Rel. 16, Last annotation update)
DE Hemolysin precursor.

GN SHLA.
 OS Serratia marcescens.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Serratia.
 OX NCBI_TaxID=615;
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 31-40.
 RC STRAIN=SNB;
 RX MEDLINE=88257037; PubMed=3290200;
 RA Poole K., Schiedel E., Braun V.;
 RT "Molecular characterization of the hemolysin determinant of Serratia
 marcescens";
 RL J. Bacteriol. 170:3177-3188(1988).
 CC -1- FUNCTION: BACTERIAL HEMOLYSINS ARE EXOTOXINS THAT ATTACK BLOOD
 CC CELL MEMBRANES AND CAUSE CELL RUPTURE BY MECHANISMS NOT CLEARLY
 CC DEFINED.
 CC -1- FUNCTION: CELL-BOUND HEMOLYSIN, WHICH RELEASES HEME-IRON FROM
 CC ERYTHROCYTES BY INTERACTION WITH THE ERYTHROCYTE MEMBRANE. SHLA
 CC REQUIRES SHUB FUNCTION.
 CC -1- SUBCELLULAR LOCATION: Outer membrane.
 CC -1- SIMILARITY: TO P. MIRABILIS HEMOLYSIN (HPMA).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC
 DR EMBL; M22618; AAA50323.1; -.
 DR PIR; A28182; A28182.
 KM Hemolysis; Toxin; Outer membrane; Signal.
 FT SIGNAL 1 30
 FT CHAIN 31 1608 HEMOLYSIN
 SQ SEQUENCE 1608 AA; 165078 MW; D669B476FE7DAD51 CRC64;

alignment_scores:
 Quality: 213.00 Length: 1609
 Ratio: 0.282 Gaps: 71
 Percent Similarity: 46.986 Percent Identity: 18.148

alignment_block:
 US-09-303-518D-653 x HLYA_SERMA ..
 Align seg 1/1 to: HLYA_SERMA from: 1 to: 1608

```

172 GCGGAAATTAAGCAAGTTGTCAGTCGCGGCGAAGATATTGAGCTTTA 221
|||||
23 AAlaAspAlaGlyAlaTyrAla.....AlaGluIleValAlaAlaAsp 37
222 CACAAAAAAGGAGGAGTGGCGCAATCGATGCAAGAAAGCCCGCATGA 271
::: ||||| |||||::: ||| |||:::
37 nGlyAlaAsnGlyProGlyValSerThrAlaAlaThrGlyAlaGlnVal 54
272 TTGATTTTCTGTGATCGCGTAAGCGCGTGGCGCATTTGGCGGCGCAT 321
::|||::: |||::: ||||| |||||:::
54 aAlaSpIleValAlaProAsnGlyAsnGly.....LeuSerHisAsn 67
322 CAATAT.....ATTGGAGCGCTGGC 341
|||||
68 GlnTyrGlnAspPheAsnValAsnGlnProGlyAlaValIleuAsnAsn 84
342 ACATTAAGCGCGCTATTAACATGTTGATTTGGTCGGAGGAGCAACATC 391
::: |||||::: ||||| |||||:::
84 rArgGlnAlaGlyLeuSerGlnLeu...AlaGlyIleuGlnValAlaAsn 100
392 CCGATCAGCAGCCGCTTTCTTACCAATTTGCAAAAGAAATATATATATA 441
|||::: |||||::: ||||| |||||:::
100 rAsnIleuGlnGlyIleuValAlaSerValIleu.....AsnGlnVal 114
442 GCAGGAGCTAACGGCCATCCTATGCGCGCATATATCATATGCGCGCTT 491

```

```

115 ILeGlyArgAsnProSerIleuHisGlyGlnGlnIle..... 128
492 GCACAAATTTTGACAGATGCAGAACCTGTGTGAGATGACCACTATTATG 541
128 ..... 128
542 ATGGTGAAATACGCTGATTAAATTAATACCGTATCGTGTGCAATC 591
129 .....Phe 129
592 GGAGCAGCAGACAAATATTGGCGGTCTGATGAAGACGAACCAAT..... 636
|||::: |||||::: ||||| |||||:::
130 GlyMetAlaAlaAspTyrValIleuAlaAsn.....ProAsnGlyIle 143
637 ..AACCGGAAAGTTTCATATCATATGTCAGACGCGATTTCTGGCTGCG 685
143 eSerCysGlnSerCysGlyPheIleAsnThrSerHisSerIleuValI 160
686 GTGCAATACCTTTGCACAAAATGATCAGTGGTGGCAGCTCAACTTA 735
||| |||::: ||||| |||||:::
160 aGlyAsnProIleuValGlnAsnGlyValIleuGlnGlyIleuSerThrPhe 176
736 GTTACGCAAAAAATTAACATAGCCCATATGTTTTCACCAAGGAGG 785
|||::: |||||::: ||||| |||||:::
177 GlyAsnArgAsnThrIleu...SerIleuAsnGlyThrIleuAsnAlaGly 192
786 CTCATTGGC.....GACAGTGGCTGCACCAATGTTA 817
||| |||::: ||||| |||||:::
192 yValIleuAspIleuIleAlaProIleAspSerArgGlyGlnValIleu 209
818 TCTATGATGCCCAAAAGCAAAAGTGTATTAATGAGGTATTTGCAACA 867
::: |||::: ||||| |||||:::
209 aGlnAspPheGlnSer.....AsnGly.....LysVal 219
868 GGCACCCCTATATAGAAAAAAGCAATGGCTTCACGATGTTGCTGAAGA 917
::: |||||::: ||||| |||||:::
220 ThrSerAlaAlaIleAsnAlaIleSerGlyIleuAsnArgValAlaArg 236
918 TTGGTTCTATGATGAATCTTTGCTGAGATACCAATCAGTATCTTACG 967
||| |||::: ||||| |||||:::
236 p.....GlyThrValGlnAlaSerGlnGln 245
968 AACCAATCAAAATGGGAATACTTTTACCAACATAATATATGCG... 1014
||| |||::: ||||| |||||:::
245 eProThrAlaIleuAspSerTyrTyrIleuGlySerMetGlnAlaGlyArg 261
1015 .....GCAGGAAA 1022
262 IleAsnIleIleAsnThrAlaGlnGlySerGlyValIleuAlaGlySer 278
1023 AATGATGCCAAACATTAACATATTCTTACTTATATGATTAATAAACAC 1072
::: |||||::: ||||| |||||:::
278 rIleuAsnAlaGlyAspIleuLeuValIleuAlaTyrAspIleuArgSer 295
1073 GAACCGTTCAATGTTAATGTTCTTTTACGAGACAGACAGAACGCT 1122
::: |||||::: ||||| |||||:::
295 IuSer...ArgValAspAspAlaSerSerAsnLysAsnGlyLysAsn 310
1123 GTTATCATGCTCAGGTGGGCTC.....AACAGTATATGACCC 1160
||| |||::: ||||| |||||:::
311 TyrGlnAsnTyrArgGlyGlyIleTyrValAsnAspArgSerSerSerG 327
1161 CACAGCTGAAT.....AATGAGCAAAATATTCTTATATGACA 1198
|||::: |||||::: ||||| |||||:::
327 nThrLeuThrArgThrGluLeuLysGlyLysAsnIleSerIleuValAla 344
1199 AAGGAAGAAGGATGATGATCTTACGACGAACATCAACCAAGCGCGGC 1248
::: |||||::: ||||| |||||:::
344 sPAsnHisAlaHis.....LeuThrAlaThrAspIleuArgGlyLysAsp 358
1249 GCTTTGATTATTTGAGGTAATTTTACGCTCTGCTTAAAAACAGCAAAC 1298
|||::: |||||::: ||||| |||||:::

```

```
359 IleThrLeuGlnGlyGlyLysLeuThrLeuAspGlyGlnGlnLeuLysGly 375
1299 GTGCGAAGGCGCGCGCTTCATATCATGATGGCAATCCGTTACTTGA 1348
375 nThrGlnGly.....HisthrAspAspArgTrpPheTyrSerTrpG 389
1349 AAGTAAAGCGGTGGCAACGACCGCTGTCCAAATCGGCAAGGACG 1398
389 InTyrAsp...ValThrArgGlnArg...GlnGlnLeuGlnGlnGln 403
1399 CTGCTGGTTCAGCCCAAGGGGAAACCAAGGCTCGCTCGCTGGCGCA 1448
404 SerThrValAlaIleSerGlySerAlaLysLeuIleSerThrGlnGln 420
1449 CGGTAAAGTCATC...TTAGATCAGCGGACGAGATCAAGCAAAAC 1495
420 PValLysLeuLeuGlyAlaIleAsnValSerAlaAspArgAlaLeuSerVal 437
1496 AAGCCTTTACTGAATCGCTGTGGTCAAGCGGAGGAGCGTCAACTG 1545
437 ysaIalaIaArgAspValHisLeuAla.....GlyLeuValGlnLys 450
1546 AATGCCGATATCATGTTCAACCCCGAACACTATTTCGGCTTGGCG 1595
451 AspLysSerSerGlu.....ArgGly 457
1596 CGGACGTTTGGATTGTGACGGGATTCGCTTTCACCGCATTCGAA 1645
457 YTYrGlnArgAsn...HisThrSerSerLeuArgThrGlyArgTrpSer 473
1646 ATACCGATGAGGGCGGATGATTTGTACACCAACCAAGCAAGAAC 1695
473 smSerAspGlnSerGlnSerLeuLysAlaSerGlnLeuArgSerGln 489
1696 ACCGTTACCATTTACAGCAATTAAGATATTACACACCGGCG... 1737
490 GlnLeuThrLeuLysAlaGlyArgAsnValSerThrGlnGlyAlaLysVal 506
1738 .....AATACACAACTTGGATGACAAAAGAAATGGCTTACA 1777
506 HisAlaGlnArgAspLeuThrIleAspAlaAspAsnGlnIleGlnValG 523
1778 AC..... 1779
523 LysAlaGlnLysThrAlaAsnAlaLysAlaValArgAspLysThrSer 539
1780 ...GGTTGGTTGGCGAGAAAGATGCAACCAAGCAAGCGGCGCTCAA 1826
540 TrpGlyGlyIleGlyGlyLysAspAsnLysAsnSerAsnArgArgGly 556
1827 TCTGAATTTACCAACCGGAAGAAAGCGGATTCGACTTACTGCTTTCGGCG 1876
556 vIleSerHisAlaSerGlu.....LeuThrSerGlyG 567
1877 GAACA...AATTTAAAGGCG.....AATATCCCAACCAAGCGCAAA 1917
567 LysThrLeuArgLysAsnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 581
1918 CTGTTTTTCAGCGGACGACGACGACGCGCTACATCATTTTGAAG 1967
581 ..... 581
1968 CGGCGTGCCTCAAAATGGAAGTATCCCAAGAGAAATCGTGGGACA 2017
582 .....SerLysAlaArgGlyGlnLysGlyGlyGlnValThr..... 593
2018 ACGATTGGATGACGCGCACATTTAAAGGGAACCTTCATATTCAGGCG 2067
594 .....AlaThrHisGly 597
2068 GGACACGCGGTG.....GTTTCCGCAATGTGGCAAGTGAAG 2108
598 GlyLeuArgIleAspAsnAlaLeuSerThrThrValAspLysIleAspAl 614

2109 CGATTGGCATTTAAGCAATACGCCCAACGACCTTTTCGTTGCGA... 2154
614 a.....ArgThrGlyThrAlaPheAsnIleThrSer 625
2155 .....CCGCATCAAAAGCCACACATCTGTACACGT 2184
625 erSerHisLysAlaAspAsnSerTyrGlnSerSerThrAlaSerGlnLeu 641
2185 TCGAGTCGACGGCTGTGACAGTTGTACCCGAAAAACCATTAACGACA 2234
642 LysSerAspThrAsnLeuThrLeuValSerHisLys.....AspAl 655
2235 TAAAGTATTGCTTCA.....TTGAGCAAGACG 2263
655 aAspValIleGlySerGlnValAlaSerGlyGlyLeuSerValGln 672
2264 ACATCAGAGCGCAATGTCAGCCTT...GCCGATCAGCTCATTTAATCTC 2310
672 erLysThrGlyAsnIleAsnValLysAlaAlaGlnArgGlnGlnAsnIle 688
2311 .....ACAGACTTGCACACTCAACGCGCATCTTAGTCAGG 2348
689 AspGlnGlnLysThrAlaLeu...ThrValAsnGlyTyrAlaLysGlnAl 704
2349 CGGACACACGACACTTACGTT.....ACGCGCAAG 2380
704 aGlyAspLysGlnTyrArgAlaGlyLeuArgIleGlnHisThrArgAsp 721
2381 CC.....ACCCAAAGCGCAACCTCAGCCTCGTGGCGAAT 2415
721 erGlnLysThrThrArgThrGlnAsnSerAlaSerSerLeuSerGlyGly 737
2416 GCC.....CAAGCAACATTTAATCAAGCCACATTAACGGCAACATC 2459
738 SerValLysLeuLysAlaGlnLysAspValThrPheSerGlySerLysLe 754
2460 GGCCTTCGACAAATGCTTCATTTAATCTAAGCAACACCGCTACAAACG 2509
754 uValAlaAspLysGlnLysAlaSerValSerGlyAsnLysVal.....S 769
2510 GCAGCTGACGCTTTCGACACACGCTAAGCGCAACGTAAGCATTTGCCA 2559
769 erPheLeuAlaAlaAspAspLysThrAlaSerAsnThrGlnThrLys 785
2560 CTCACAGGCAATGTCCTCCCTAGCCGATPAAGCAGTATTCATTTGAAA 2609
786 IleGlyGlyGly.....PheTyrTyrThrGln 794
2610 C.....AGCCGCTTACCGGAAAAATCAGCGGCGCAAGATACGCAT 2653
794 yGlyIleAspLysLeuLysSerGlyValGlnAlaGlyTyrGlnAsnL 811
2654 TACACTTAAAGACAGCGAATGACGCTGCCGTGCGGACGGAATTA... 2700
811 yThrGlnAlaGlnSerSerLysAlaIleThrSerGlySerAspValLys 827
2701 GCGAATTTAAACCTTGAC.....AAGCGCAC 2726
828 GlnAsnLeuThrIleAsnAlaArgAspLysLeuThrGlnGlnGlnAla 844
2727 CATTAACCTCAATTCGCGCTATGACACGATGAGGAGGCGGCAACCG 2776
844 nHisSerValGlyGlyAlaTyrGlnGlnAsnAlaAlaGlyValAspHisL 861
2777 GCAGTGCGGACGATGCGCGCGCGCGCTGCGGCGCTCCATATATCC 2826
861 euAlaAlaIaAspThrAlaSerThrThrThrLysThrAspValGly 877
2827 GTTACCGCGCACTTCGCGCAGATCCGTTTCAACACGCTGACGGTAA 2876
878 ValAsnIleGlyAlaAsnValAspTyrSerAlaValThrArgProValGln 894
```

```

2877 C.....GCCAATTGAACGCTCAGGAGACATTCGGCT 2908
      :|||||:.....:|||||:
894 uargAlaValGlyLysAlaIaLysLeuAspAlaThrGlyValIleAsn 911
2909 TTATGCTGAATC.....TTCGGCTACCG 2934
      :|||||:.....:|||||:
911 spIleGlyIleGlyAlaProAsnValGlyLeuAspIleGlyAlaGln 927
2935 AGCGCAAAATTGAAGTGGCGGAAAGTTCCGAGAGCACTTACACCTGGC 2984
      :|||||:.....:|||||:
928 GlyIleSerSerGlyLysArgSerSerSerGlnAlaValIleSer 944
2985 TGTCACAAATCCGGCAACGACCCGTAAGCTCGAGCAATGTACGGTAG 3034
      :|||||:.....:|||||:
944 IVal...GlnAlaGlySerIleAspIleAsnAlaLysGlyGlnValArg 960
3035 TGAGAGAAAGACACACACCGCTGCCAATTTAATTTACCCCTG 3084
      :|||||:.....:|||||:
960 spGlnGlyThrGlnThrGlnAlaSerLysGlyAlaValAsnLeuThrAla 976
3085 CAAGACACACGCTGATGCCGGCGCATGGGCTTATCAGCTTATCCGCA 3134
      :|||||:.....:|||||:
977 AsperHisArgSerGlnAlaIaIaAsnArg..... 987
3135 AGACGGCAGTTCCGCTGCATATCCGTCACAGAACAGACGCTTCCG 3184
      :|||||:.....:|||||:
988 .....GlnAspGlnIleSerArg 994
3185 ACMAACTGGCAGAGCGGAGAAACAGAGCGCTTGACGGCAAAACAG 3234
      :|||||:.....:|||||:
994 spThrArgGlySerAlaGlyValArgValTyrThrThrGlySerAsp 1010
3235 GCACACTGGCGCCCAA.....CAACAGCGGCAAAAGA 3269
      :|||||:.....:|||||:
1011 LeuThrValAspAlaLysGlyGlnGlyThrGlnArgSerAsnSer 1027
3270 CAACCGCA.....AGCCTGACGCGCTGATTCGGCGCGGC 3307
      :|||||:.....:|||||:
1027 rAlaSerGlnAlaValThrGlySerIleAspAla.....AlaAsnGly 1042
3308 GCATGCCACCGAAAGCAGAAATGTTGCCGAACCGCGCGCAGGCA 3357
      :|||||:.....:|||||:
1042 IeAsnValAsnValLysLysAspAlaIleTyrGlnGlyThrAlaLeuAsn 1058
3358 GCGCGGAAATCCGCGCATTTATCAGCGGAGAGAAAGAAACGCGT 3407
      :|||||:.....:|||||:
1059 GlyGlyArgGly.....LysThrAlaVal 1066
3408 GCAGCGGATTAAGACACCGCTTG.....GCGAAACAGCGCGAAGCG 3451
      :|||||:.....:|||||:
1066 IAsnAlaGlyGlyAspIleArgLeuAspAlaSerAspLysGlnSer 1083
3452 AAACCGCGCGGCTACACCGCTTCCCGCGCGCGCGCGCGCG 3501
      :|||||:.....:|||||:
1083 IuSerArgSerGlyPheAsn..... 1089
3502 GATTTCGCCAACCGAGCCCAACCGCAACCCCAACGCGAGCGACT 3551
      :|||||:.....:|||||:
1089 ..... 1089
3552 GATCAGCCCTTATGCCAATAGCGGTTTGATTTCCGCGCAGCTCA 3601
      :|||||:.....:|||||:
1090 ValLysAlaSerAlaLysGlyIlePheThrAlaAspSerLysAsnPheG 1106
3602 ACAGCGTTTCCGCGTACAGAGCAATGTGACCGCGCTTTCGCGAAGC 3651
      :|||||:.....:|||||:
1106 IValIaGlyPheGlyGly.....ThrHisAsn 1115
3652 CGCGGACAGCGCTTGACAAAGCGCATCCGGACACCAACACTACCG 3701
      :|||||:.....:|||||:
1116 GlyLysSerSerSerThrAlaGlnValGlyAsnIleSerGlyGlnI 1132
3702 TTCGCAAGATTTCCGCGCTACCGC.....CAACAACGCGAC 3739

```

```

1132 nGlyValGlyLeuLysAlaGlyArgAspLeuThrLeuGlnGlyThrAspV 1149
      :|||:.....:|||||:
3740 TCGCCCAATTCGT..... 3753
1149 AllysSerGlnGlyAspValSerLeuSerAlaGlyAsnLysValAlaLeu 1165
      :|||:.....:|||||:
3754 .....ATCGCAAAAA 3764
1166 GlnAlaIaGlnSerThrGlnThrArgLysGlnSerLysLeuSerGlyAs 1182
      :|||:.....:|||||:
3765 CCTCGGACGCGCGCGCTCGCATCTGTTTTCGACACAGCGAGCGAA 3814
      :|||:.....:|||||:
1182 nIleAspLeuGly...AlaGlySerSerAspSerLysGlnGlyThrGlyG 1198
3815 ACACCTTCGACGACGCG.....ATCGCAACTCG 3843
      :|||||:.....:|||||:
1198 LysAsnLeuSerAlaGlyGlyAlaPheAspIleAlaLysValAsnIleu 1214
3844 GCAGCGCTTGCACAGCGTCCGTTTTCGGCAATACGCGATCGCAGCTT 3893
      :|||||:.....:|||||:
1215 AlaThrGluArgGlnGlyAlaThrIleAlaSerAspGly..... 1227
3894 CGACATCGGCATCAGCGCG..... 3912
1228 LysValThrLeuSerAlaAsnGlyLysGlyAspAlaLeuHisLeuG 1244
      :|||:.....:|||||:
3913 ..GCGCGGCTTATTCAGCGCGAGC..... 3936
1244 IlnGlyAlaLysValSerGlyGlySerAlaAlaLeuGlnAlaLysasnGly 1260
3937 .....CTTTCAGCGCATCAGCAGCAAAATCCCGCGCGCTGTGCA 3980
      :|||:.....:|||||:
1261 GlyIleLeuLeuGlnSerAlaLysAsnGlnGlnHisLysAspAsnTrpSe 1277
3981 TTACGGCATTCAGCAAGATATCCGCGCAGT.....TTCG 4015
      :|||||:.....:|||||:
1277 rLeuGlyIleLysAlaAsnAlaLysGlyGlnThrPheAsnLysAspAla 1294
4016 GCGATTCGCGCATCAGACCGCACATCGCGCAACGCGCTATTCGTC... 4062
      :|||||:.....:|||||:
1294 IagGlyGlyLysValAspProAsnThrGlyLysAspThrHisThrLeuGly 1310
4063 .....CAAAAAGCGGATTACCGATACGAAAA 4088
1311 AlagIleuLysValGlyValGlnGlnAspLysThrHisAlaAs 1327
4089 GGTCAATATCGCACCCCGCGCTTGATTCACCGC..... 4125
      :|||||:.....:|||||:
1327 nThrGlyIleThrAlaGlyAspValThrLeuAsnSerGlyLysAspThrAla 1344
4125 ..... 4125
1344 rGlyLeuAlaGlyAlaArgValAspAlaAspSerValGlnGlyLysValGly 1360
4126 .....TACCGCGCGCGCATTAAGGC 4145
1361 GlyAspLeuHisValGlnSerArgLysAspValGlnAsnIleValLysVal 1377
4146 AGAT.....TATTCATTCAACCGCGCGCAACACATT 4177
      :|||:.....:|||||:
1377 IAspValAspAlaGlyLeuSerHisSerAsnAspProGlySer.....S 1392
4178 CCATCAGCGCTTATTTAGCTGTCCATATCCGATGCGCGCTTCGCGCAA 4227
      :|||||:.....:|||||:
1392 erIleThrSerLysLeuSerLysValGlyThrProArgTyrAlaIaGlyLys 1408
4228 GTCGACAGCGCGTCATATACGCGCTA 4254
1409 ValLysGlyLysLeuGlnAlaGlyVal 1417
seq_name: SwissProt_40: PMP6_CHLPN

```

seq documentation block:
 ID PMP6_CHLIPN STANDARD: PRT; 1276 AA.
 AC 092899; 09JRM2;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Probable outer membrane protein pmp6 precursor (Polymorphic membrane protein 6).
 GN PMP6 OR CP00444 OR CP0309.
 OS Chlamydia pneumoniae (Chlamydia pneumoniae).
 OC Bacteria; Chlamydiales; Chlamydiae; Chlamydia.
 OX NCBI_Taxid=83558;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CWL029;
 RX MEDLINE=99206606; PubMed=10192388;
 RA Kalman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,
 RA Olinger L., Grimwood J., Davis R.W., Stephens R.S.;
 RT "Comparative genomes of Chlamydia pneumoniae and C. trachomatis.";
 RL Nat. Genet. 21:385-389(1999).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-AR39;
 RX MEDLINE=20150255; PubMed=10684935;
 RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
 RA White O., Hickey E.R., Peterson J., Utterback T., Berry K., Baas S.,
 RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
 RA Gwinn M., Nelson W., Deboy R., Kolonay J., McClarty G., Salzberg S.L.,
 RA Eisen J., Fraser C.M.;
 RT "Genome sequences of Chlamydia trachomatis Mohn and Chlamydia pneumoniae AR39.";
 RL Nucleic Acids Res. 28:1397-1406(2000).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-J138;
 RX MEDLINE=20330349; PubMed=10871362;
 RA Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,
 RA Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;
 RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138 from Japan and CWL029 from USA.";
 RL Nucleic Acids Res. 28:2311-2314(2000).
 CC -1 SUBCELLULAR LOCATION: CELL WALL SURFACE (ELEMENTARY BODIES)
 CC -1 (POTENTIAL).
 CC -1 SIMILARITY: BELONGS TO THE PMP OUTER MEMBRANE PROTEIN FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/or-send-an-email-to-licensing@isb-sib.ch>).
 CC EMBL; AE001627; AAD18588.1;
 CC EMBL; AE002193; AAF38166.1;
 CC EMBL; AP002546; BAA98652.1;
 CC PDB; 2DPAGE; Q9Z899;
 CC TIGR; CP0309;
 CC InterPro: IPR003368; DUF145.
 CC InterPro: IPR003357; OMP.
 CC Pfam: PF02415; DUF145; 2.
 CC Pfam: PF02385; OMP; 1.
 CC Outer membrane; Signal; Multigene family; Complete proteome.
 CC SIGNAL
 CC CHAIN 24 1276
 CC CONFLICT 421 421
 CC E -> ECGGAILAFIDSGVSTKSTINNOEVLSTNA
 CC AIVSGAIVATFKCTIGNGSITDGNAGTSCGATSTED
 CC FTLLGSGTGTFTSTNTAKTGCALYSKNNLSGNTINLTSFSG
 CC NKATGPNSSANDE (IN REF. 1).
 CC SEQUENCE 1276 AA; 132127 MW; C97FA40662C88E0B CRC64;

alignment_scores:

Quality: 209.00 Length: 797
 Ratio: 0.563 Gaps: 36
 Percent Similarity: 46.530 Percent Identity: 20.828
 alignment_block:
 US-09-303-518D-653 x PMP6_CHLIPN
 Align seg 1/1 to: PMP6_CHLIPN from: 1 to: 1276
 1570 GACAACTCTATTTCGGCTTCGCGGACGCTTGGATTGTAACGGG.. 1617
 ||||| ||| ||||| ||||| ||||| ||||| ||||| |||||
 75 AsplysSerCysPheThrAsnThrGlyAlaLeuSerPheValGlyAl 91
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 1618 ...CATTCGCTTTCCTCCACCGCATTCAAATATCCGATGAAGGGCGA 1663
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 91 AsplysSerCysPheThrAsnThrGlyAlaLeuSerPheValGlyAla 108
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 1664 TGATGTCAACGCAAT..... 1680
 ||| ||| |||
 108 lAlaAsnAsnThrAsnThrAlaLeuSerPheSerGlyPheSerLeu 124

 1680 1680
 125 LeuIleAspSerAlaProAlaThrGlyThrSerGlyGlyAlaAl 141

 1681CAAGCAAGAATCCACCGCTTACCTTACGACGATAAAG 1720
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 141 eCysValThrAsnThrGlyGlyThrAlaThrPheThrAspAsnAla 158
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 1721 ATATTACTACACCGCGCAATACCAACTGGATGATGCAAAAGAAATT 1770
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 158 ervaThrLeuGlnLysAsnThrSerGlyLysAspGlyAlaAlaValSer 174
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 1771 GCCTTCAAC.....GGTTGGTTGGCGA 1793
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 175 AlatySerIleAspLeuAlaLysThrThrAlaAlaLeuAspG 191

 1794 GAAGATCCCAACCAAGCAAGCGGGCGCTCAATCTGATTTACCAACCGG 1843
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 191 nasnThrSerThrLysAsnGlyGly..... 199

 1844 AAGAAAGCGATCGACTTACTGCTTCGCGGCAACAATTAAAGCGC 1893
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 200AlaLeuSerThrAlaAsnThrThrValGlnGly 211

 1894 AATATACGCAACAACGCAACTGTTTTCGCGGACCGACACCC 1943
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 212 Asn.....SerGlyThrValThrPheSerSerAsnThr..... 222

 1944 GCAGCGCTACATCATTTAGAACCGGCGGTGTCAAAATGGAAGGTATCC 1993
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 223 ...AlaThrAspLysGlyGlyGlyIleThrLysGlyLys..... 235

 1994 CACAAGGAGAATCGTGTGGCAACGATTGATGACCGACATTTAA 2043
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 236AspSerThrLeuAsp 240

 2044 GCGGAAAACCTTCATTTACAGGCGGACAGCGGTGTTCCCGCAATGT 2093
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 241 AlaAsnThr.....GlyValValThrPheLysSerAsnThr 252

 2094 TGCCAAAGTGAGGCGATTGGCAT.....TTAAGCA 2125
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 252 fAlaLysThrGlyAlaThrPheSerSerAspAsnLeuAlaLeuThrG 269
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 2126 ATCAAGCCCAAGCAGTTTC.....GGTGGCGACCG 2157
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 269 LysAsnThrGlnValLeuPheGlnGlnLysAsnLysThrThrGlySerAlaAla 285

 2158 CATCAAGCCAC.....ACAATCGT..... 2178
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 286 GlnAlaAsnAsnProGlnGlyLysGlyAlaAlaIleCysCysTyrLeuAl 302

```

2179 .ACAGCTCGACGTGACGGGTGTGACA.....A 2206
    ||| :||| |||||:|
302 aThrIaThrAspLysThrGlyLeuAlaIleSerGlnAsnGlnGlnMetS 319
2207 GTTGTACGGAAAAACCATTCACGACGATAAAGTGATTGCTTCATTGAGC 2236
    || :|||:| |||
319 erPheThrSerAsnThrThrIaAsnGlyGlyAlaIleTyraIaThr 335
2257 MAGACGACATCAGACGATGTCAGCTGCG..... 2289
    ||| :|||:| |||||:|
336 LysCysThrLeuAspLysAsnThrThrLeuThrPheAspGlnAsnThrAl 352
2290 .....GATCAGCGTC 2299
    ||| :|||
332 aThrIaGlyCysGlyGlyAlaIleTyrThrGlnThrGlnAspPheSerL 369
2300 ATTAAATCTCAGAGACTGGCCACACTGCACGGCAATCTTAGTCAGGC 2349
    :||| :|||:| |||
369 euLysGlySerThrGlyThrValThrPheSerThrAsnThrAlaLysThr 385
2350 GGAGACAGCCACTATACGGTTACGGCAACGCCACCAAAAGCGCAACT 2399
    ||| :|||:| |||||:|
386 GlyGlyAlaLeuTyrSer.....LysGlyAsnSe 395
2400 CAGCCTCGTGGCAATGCCCAAGCAACATTT.....AATCAAGCCAGAT 2443
    |||||:| ||| :|||:| |||||:|
395 rSerLeuThrGlyAsnThrAsnLeuThrPheSerGlyAsnLysAlaThrG 412
2444 TAAAGCGCAACATCGCTTGCGAATGCT..... 2475
    :|||:| |||||:| |||||:|
412 LyrProSerAsnSerIaAsnGlnGlnGlyCysGlyAlaIleLeu 428
2476 TCATTAAATCTAAGCAACACCCGTACAAAGCGCAGCTGACGCTTC 2525
    |||||:| ||| :|||:| |||||:|
429 SerPheLeuGlnSerIaSerValSerThrLysGlyLeuThrLleG 445
2526 CGACAAAGCTAAGCAACAGTAAAGC.....CATTCGCGACTCAAGGCA 2569
    :|||:| |||||:| |||||:|
445 uAspAsnGlnAsnValSerLeuSerGlyAsnThrIaThrValSerGly 462
2570 ATGTCCTCCGTAGCCGATAGGCAATATTCAT..... 2601
    :|||:| |||||:| |||||:|
462 LysAlaIleTyraIaThrLysCysAlaLeuIleGlyAsnThrThrLeuThr 478
2602 TTGTAAACAGACCGCTTTACCGAAAAATCGAGCGCGCAAGATACGGC 2651
    |||||:| ||| :|||:| |||||:|
479 PheAspGlyAsn.....ThrIaGlnThrIaGlyGly.....Al 490
2652 ATTACACTTAAAGACAGCGAATGAGCGCTGCGCGGCGACGGAATTAG 2701
    |||||:| ||| :|||:| |||||:|
490 aLeuTyThrGlnThrGlnAspPheThrLeuThrGlySerThr.....G 505
2702 GCAATTAAACCTTGACAAAGCCACATTCACATC..... 2736
    |||||:| |||||:| |||||:|
505 LysThrValThrPheSerThrAsnThrAlaLysThrAlaGlyAlaLeuHis 521
2737 .....AATTCGCGCTATCGACAGATGCGGAGCGCGCAACCGG 2777
    |||||:| ||| :|||:| |||||:|
522 ThrLysGlyAsnThrSerPheThrLysAsnLysAlaLeuValPheSerG 538
2778 CAGTGGCGGAGATGCGCGCGCGCGCTTCCGCGCTTCATTATTCG 2827
    :|||:| ||| :|||:| |||||:|
538 LysAsnSerAlaThrAla.....Thra 545
2828 TTACGCGCGCACTTCGCGAATCCCGTTTCAACAGCGTGAAGTAAC 2877
    :|||:| |||||:| |||||:|
545 IeThrThrThrAspGlnGlnGlyCysGlyAlaIleLeuLysAsn 561
2878 GGCAATTAAGCGTACGAGACATTCGCTTATGTCGAACTCTCG 2927
561 ..... 561
2928 CTACCGCAGCGCAATTGAAGCTGCGGAAAGTTCGAAAGCACTTACA 2977

```

```

562 .....:| |||||:| |||||:|
2978 CTTGGCTGTCAACATACCGGCAAGAACCCGTAGCTCCGACGAAATG 3027
    :|||:| ||| :|||:| |||||:|
571 erLeuThrLeu.....ThrGlnAsnGlnSerLeuSerPheIleAsnAsn 585
3028 ACG.....GTAGTGA 3038
    ||| :|||
586 ThrAlaLysArgSerGlyGlyIleTyraIaProLysCysValIleSe 602
3039 AGCAAAAGACAAACACACCGCTGCGAAAATCTTAATTTCACCTGCAA 3088
    :||| :|||:| |||||:|
602 rGly.....SerGlnSerIleAsnPheAspGlyAsn 613
3089 ACGACACGTCGATGCGCGGCATGGGTTATGACTTATCCGCAAGAC 3138
    :|||:| |||||:| |||||:|
613 hrIaGlnThrSerGlyGlyAla.....IleTySerLysAsn 625
3139 GCGGAGTTCGCGCTGCATAATCCGTCGAAGAACAGAGCTTTCGACAA 3188
    :|||:| |||||:| |||||:|
626 LeuSerIleThrIaAsnGlyProValSerPheThrAsnAsnSerGly.. 641
3189 ACTCGGCAAGCGGGA.....GAACGAGAGCGCGCT 3220
    |||||:| ||| :|||:| |||||:|
642 ...GlyLysGlyAlaIleTyrIleAlaAspSerGlyLeuLeuSerL 657
3221 TGACGGCAAAACAGGCAACTTGGCGCCCAACACAGCGGCGGAAAAAGAC 3270
    ||| :||| :|||:| |||||:|
657 euGlnAlaIleAspGlyAspIleThrPheSerGlyAsnArgAlaThrGln 673
3271 AAGCGCAAAAGCCTTACGCGCTG.....ATTGCGCGCGCGCAANTGCCAC 3317
    :|||:| ||| :|||:| |||||:|
674 GlyThrSerThrProAsnSerIleHisLeuGlyAlaGlyAlaLysIleThr 690
3318 CGAAAGCGCAAAAGCTTGCC.....GAGCGGCGCC 3349
    :|||:| ||| :|||:| |||||:|
690 rLysLeuAlaIaIaProGlyHisThrIleTyrPheTyraSProler 707
3350 GGCAGCGCAGCGGGGAAATGCGCGCATTCAGAGCGGAGGAAGAA 3399
    :|||:| ||| :|||:| |||||:|
707 hrMetGlnAlaProIaSerGlyGlyThrLleGlnGluLeuValIleAsn 723
3400 AAGCGGTGCAAGCGGATTAAGACACCGCTTGCGGAACAGCGCGAAGC 3449
    |||||:| ||| :|||:| |||||:|
724 ProValValLysAlaIleValProPro.....GlnP 735
3450 GGAACCGCGCGGCTACACCGCTTCCCGCGCGCGCGCGCGCGC 3499
    ||| :||| :|||:| |||||:|
735 rLysAsnGlyProIleAlaSerValProValValProValAlaProAla 751
3500 GGATTTGCCGCAACCGCACCCCAACCGCAACCCCAAC 3538
    |||||:| |||||:|
752 .....AsnProAsn 754

seq_name: SwissProt_40:N100_YEAST
seq_documentation_block:
ID N100_YEAST STANDARD: PRT: 959 AA.
AC 002629;
DT 01-JUL-1993 (rel. 26, Created)
DT 01-JUL-1993 (rel. 26, Last sequence update)
DT 01-OCT-1996 (rel. 34, Last annotation update)
DE Nucleoporin NUP100/NSP100 (Nuclear pore protein NUP100/NSP100).
GN NUP100 OR NSP100 OR YKL068W OR YKL336.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93054906; PubMed=1385442;
RA Wente S.R., Rout M.P., Blobel G.;

```


349	lnAsnSerGIthrSerIySerSerGIyLeuPheGlyGlnAsnAsnGln	365
1684	GACAAAGATACCCCGTTACCATTAACAGGCAATTAAGATATTACTACAC	1733
366	LysGlnProAsnThrPheThr...GlnSerAsnThrGlyIleGlyLeuPhe	381
1734	CGGCAATTAACAACAACCTTGATAGCAAAAAAGAAATTCCTACACGGTT	1783
381	eGIyGlnAsnAsnGlnGlnGlnGlnSer.....ThrIleL	394
1784	GGTTTGGCAGAAAGATGCACCAACAAAACAGGCGGCTC.....	1824
394	eupheGlyAlaIySProAla...GlyThrThrGlySerLeuPheGlyGly	409
1825	AATCTGAATTACCAACCGGAGAGAGCGGATGCACATTACTGCTTCGG	1874
410	AsnSerSerThrGlnProAsnSer.....LeuPheG	420
1875	CGGAACCAATTTAAACGGCAATACAGCAACA.....AACGGCAAC	1918
420	YthrThrAsnValProThrSerAsnThrGlnSerGlnGlnIyAsnSerL	437
1919	TGTTT.....TTCAGCGGACAGCGACA	1941
437	eupheGlyAlaThrIySerLeuThrAsnMetProPheGlyIyAsnProThr	453
1942	CCGACGCGCTACAAATCATTTAGGAAGCGGTGGTCAAAAATGGAAGTAT	1991
454	Ala.....AsnGlnSerIySerIyAsnSerLeuPheGlyThrL	467
1992	CCCACAGGAGAAATC.....GTGTGGACACAGATTGATCGACGCA	2035
467	sProAlaSerThrThrGlySerLeuPheGlyAsnAsnThrAlaSerThr	484
2036	CATTTAAGCGGAAACTCCATATTCAGAGCGGACAAACGGGTGGTTCC	2085
484	hValProSerThrAsn...GlyLeuPheGlyAsnAsnAlaAsnAsnSer	499
2086	CGAAATGGTGGCAAAAGTGAAGGCGATTGGCATTTAAGCAATCAGGCCA	2135
500	ThrSerThrThrAsnThr.....	505
2136	AGCATTTTTCGTGTCGACCCGATCAAGCCACACAATCTGTACAGTT	2185
506	.GlyLeuPheGlyAlaIySProAspSerGlnSerIySProAlaLeuGly	522
2186	CGGACGTGAACGGGTGACAACTTGACGAAAAACATTACACGAGAT	2235
522	IyIyLeuPheGlyAsnSerAsnSerAsnSerThrIleGlyGlnAsn	538
2236	AAAGTGAATGCTTCATTGAGCAAGACCGACATCAGAGCAATGTACGCT	2285
539	LysProVal.....	541
2286	TGGCGATCACGCTATTAAATCTCAAGAGATTGCCACACTCAAGGCA	2335
541	541
2336	ATCTTAGTCAGCGGAGACACGCACTATACGTTACGCGACGCGCAC	2385
542PheGlyGlyThrThrGlnAsnThrGlyLeuPheGlyAlaThr	555
2386	CAAAAGCGCACTCAGCCTCGTGGGCAATGCCCAAGACATTAAATCA	2435
556	...GlyThrAsnSerSerAlaValIyGlySerThrGlyLysLeuPheGly	571
2436	AGGCACATTAAAGCGCAACATCGGCTTCGGACATGCTTCATTAAATC	2485
571	n.....AsnAsnAsnThrLeuAsnValGlyThrGlnAsnValProP	585
2486	TAAACAACAAGCGCGTACAAAC.....GGCAGTCTGACGCTTCC	2526

328 .GluGlyValTyrIlysAspLysProIlysAspLysProSerAsnThrIrg 344
1217 TACTTACGACCAATCAACCAAGCGCGGCTTTGTATTTGAGGGT 1266
:::||||| ||||| :::::
344 InAsnAsnAlaAsnAsnInglInAsnSerAlaGlnAsnAsnSerAsn 360
1267 AATTTTACGGCTCGCTTAAAAACAC...CAACGTGGCAAGCGCGG 1313
::: ||||| ||||| :::
361 ThrGlnValIleAsnProIleAsnSerAlaGlnIleAsnThrGlnIleGlnPr 377
1314 CGTTGATTCAGTGTAGTACGAGTACGTTACTGTGAAGTAAAGCGCGG 1363
::: ||||| ||||| :::
377 ThrGlnValIleAspGlyProIleAlaGlyIlyLysAspThrValIleVala 394
1364 CAACGACCGCTCTCCAAATTCGCAAGCGACGCTGCTGTT... 1407
|||||::: :: |||||:::
394 snIleAspArgIleAsnThrAsnAlaAspGlyThrIleLysValGly 410
1408CAACCAAGGGGAAAAACCAAGCTCGTACGCTGGCGGAC 1451
::: ||| :::: |||||
411 TyrIleAlaSerIleuThrThrAsnAlaAlaIleuHisIleGlyLysG 427
1452 TAAAGTCATCTTATGATCAGACGCGGACGATCAAGCAAAAACAGCCT 1501
|::: |||||::: |||||
427 yGlyIleAsnIleuSerAsnGlnAla..... 435
1502 TTAGTGAATCGGCTTGTGTCAGCGGAGCGGACGCTGCAACTGATGCC 1551
||| |||||
436SerGlyArgThrIleuLeuValGlnAsnIleu 445
1552 GATATTCAGTTCACCCGAC...AACTCTATTTCGGCTTCGCGGCGG 1598
::: |||||::: |||||
446 ThrGlyAsnIleThrValAspIlyProIleuArgValAsnAsnGlnIleVal 462
1599 ACGTTGATTTGAACGGGCACTTCCTTCCTCCACCGCATTCACAAATA 1648
| ||| ||| ||| ::|||:::
462 yGlyTyrAlaIleuAlaGlySerSerAlaAsnIleuIleuLysAlaGlyT 479
1649 CCGATGAAGGGCGATGTGTCAACCAATCAAGCAAAAGATCCACC 1698
|||||
479 hrAsp.....ThrLysAsnGlyThr 485
1699 GTTACCATTCAGCAATTAAGATTTACTACA..... 1731
::: |||||::: |||||
486 AlaThrPhe.....AsnAsnAspIleSerIleuGlyArgPheValAsnIle 500
1732ACCGCAATTAACCAACTTGGATAGCAAAAAG 1765
|||||::: |||||
500 ulysValAspAlaHisThrAlaAsnIleuLysGlyIleAspThrGly... 515
1766 AATTTGCTACACGCTTGTGGTGGAGAAAGATGCAACCAAAACGAC 1815
||||| |||||
516AsnGlyIlePheAsnThrIleuAspPheSerGlyValThr 528
1816 GGGCGGCTCAATCTGAATTTACCAACCGAAGAGCGGATCGCACTTACT 1865
|||||::: |||||
529 GlyLysValAsnIleAsn.....LysLe 536
1866 GCTTCCGCGGCAACA.....AATTAAACGCA 1894
::: |||||::: |||||
536 uIleThrIleAspThrAsnValAlaValLysAsnIleAsnIleAsnGln 553
1895 ATATACGCAAAACAAAGCAACTGTTTTCAGCGGACGACCGACCG 1944
::: |||||::: |||||
553 euValValIlyThrAsnGly.....ValSerValGlyLysThr 566
1945 CACGGCTCAATCATTTAGACGGGCTGTCTAAATGAAGATATGCC 1994
||| :::: |||||
567 HisPheSerGluAspIleGlySer..... 574
1995 ACAAGAGAAATCTGTGGACACAGATGTGATCGACCGCACTTTAAG 2044
|||||::: |||||
575 .GlnSerArgIle.....AsnThrValAlaGln 583
2045 CGGAAACTTCATATTACGGCGGACAAAGCGGTGTTCCGCAATTT 2094
|||::: ||| ::|||
583 euGlnThr.....GlyThrArgSerIlePheSerGlyVal 595
2095 GCCAAAGTGAAGCGCATTTGGCATTTAGCAATACGCCCAAGCAATTT 2144
::: |||||
596 LysPheLysSerGly..... 600
2145 CGGTGTGCAACCCGATCAAGCCACACAAATCTGTACGTTGAGATGGA 2194
600 600
2195 CGGGTGTGACAGTGTGTACCGAAAAACATTACGACGATTAAGTATT 2244
|||||::: |||||
601GlnLysLeuValIleAspIleuPheThr 610
2245 GCTTCATTGACAAAGCCGACATCAGAGG...AATGTCAGCCTT... 2286
::: :::: |||||
611 SerProIleAsnThrPheAspAlaArgAsnIleLysAsnValGlnIleTh 627
2287GCCATCAGCTCATTTAAATCTCACAGGACTTGCACAC 2326
|||::: |||
627 rArgLysPheAlaSerSerThrProGlnAsnProIlePglyThrSerLysL 644
2327 TCAACGGCAATCTTATGACAGCGGAGACACGCACTATACGTTACGCC 2376
|||
644 euMetPheAsn.....AsnLeuThrIleuGlyGln 653
2377 AACGCC.....ACCCAAAGCGCAACCTGACGCTGCGGC... 2412
|||||
654 AsnAlaValMetAspTyrSerGlnPheSerAsnIleuThrIleGlnLys 670
2413AATGCCAAGCAACATTTATCA... 2436
||| |||||
670 pheIleAsnAsnGlnGlyThrIleAsnThrIleuValArgGlyGln 687
2437 .GCCACATTAAAC...GCCACACATCGGCT.....TCGGACAT 2472
|||||
687 alaIleThrIleuAsnValGlyAsnAlaAlaMetPhePheSerAsn 703
2473 GCT.....TCATTTATCTAG 2489
::: |||||
704 ValAspSerAlaThrGlyPheThrGlnProIleuMetLysIleAsnSerAl 720
2490 CAACAAAGCGGCTACAAAC..... 2508
::: |||||
720 aglnAspLeuIleLysAsnLysGlnHisValIleuLysAlaLysIle 737
2509GCCAGTCTGACGCTTCCGACACGCTTAAGCAAACTTAAGC 2550
|||||
737 leGlyTyrGlyAsnValSerLeuGlyThrAsnSerIleSerAsnVal... 752
2551 CATTCGCACTCAACGCAATGTCCTCCTAGCCGATAGCAAGATTTCA 2600
|||
753AsnLeuIleGlnGlnPheLysGlyArgLeuAlaLeu 765
2601 TTTTGAAGAACGCCGCTTTACGGAAAAATCAGCGGCGGACAGATACG 2650
::: |||||::: |||||
765 rAsnAsnAsnAsnArgMet.....AspIleC 774
2651 CATTACCTTTAAAGACAGCAATGAGCGTGCCTCGGACAGCAATTA 2700
::: |||||::: |||||
774 yVal...ValArgAsnThrAspAspIleLysAlaCysGlyThrAlaIle 789
2701 GGCATTTA.....AACCTTGACAAAGCCACCATTAACACTCA 2738
|||||
790 GlyAsnGlnSerMetValAsnAsnProIleAspAspTyrLysThrIleuIleG 806
2739 TTCGCGCTATGCAAC.....GATGGGACGCGCGCA 2772
::: |||||::: |||||
806 yLysAlaIleThrLysAsnIleGlyIleSerLysThrAlaAsnGlySerLys 823


```

793 yThrserrThrrpGlyAsnAsnThrSerIleGluThrThrleuThrleuA 810
3356 CAGCGGGGAAATCCCGCATTTATGCGCGGAGAGAAAGAAAAACGG 3405
810 laAsnGlyAsnIleGlyHisIleValIleLeuGlu..... 821
3406 GTGACGGGGGATTAAGACACCGCCTTGCGAAACAGCGCGAAGCGGAAC 3455
822 .....GlyAlaIleGlnVa 825
3456 CCGGCGGCTACACCGCCTTCCCGCGCGCGCGCGCGCGCGGATT 3505
825 laSnThrThrThrThrGly..... 831
3506 TGCCCGAACCGACGCCCAACCGCAACCCGACGCGACTGATC 3555
831 ..... 831
3556 AGCCGTATGCAATAGCGGTTTGAGTGAATTTCCGCGACGCTACAG 3605
832 .....ThrTh 833
3606 CGTTTCCGCTACAGAGCAATTGGACCGCGTGTTCGCGAAGACCGCC 3655
833 rThrIleLysValGlnAspAsnAlaAsnAlaAsnPheserGlyThrGlnT 850
3656 GGAAGCGCGTTGGACAAAGCGCATCGGGACACCAACACACACGCTGTC 3705
850 hTyrThrleuIleGlnGlyAlaIleArgPheAsnGlyThrleuGlySer 866
3706 CAAGATTTCGCGCTACCGCGCAACCAACCGACCTGCGCAATCGGAT 3755
867 ProAsnPhE..AlaValThrGlySerAsnArgPheValAsnIleThrleuI 883
3756 GCAGAAAAACCTCGCGACGGCGCGCTGCGCATCTGTTTTCGCAACAC 3805
883 laArg.....AlaAlaAsnGlnAspTyrValIleThrArgThrAsn 896
3806 GGACGGAACACACTTCGACGCGGACGCGACCTCGGACGCGCTGCGC 3855
897 AsnAlaGlnAsnValValThr.....AsnAspIleAl 907
3856 CACGTCGCGTTTTCGCGCAAT.....ACGGCAT 3884
907 aAsnSerProPheGlyAlaProGlyValAlaAspGlnAsnValThrThp 924
3885 CCGCAGCTTCGACATCGCATCGCGCGCGCGCTTTTACGAG... 3931
924 heValAsnAlaThrAsnThrAlaAlaTyrAsnAsnleuLeuAlaLys 940
3932 .....GCAGCTTTCAGACGCGCATCGACGAGCAAAATCGCG 3966
941 AsnSerAlaAsnSerAlaAsnPhEValGlyAlaIleValThrAspThse 957
3967 CCGCGGCTCGCATTCAGCATTCAGCGCAATACCGCGAGTTTCGG 4016
957 rAlaAla.....IleThrAsnValGlnleuAspAlaLysAspIleG 972
4017 CCGATTTCGCGATCGAACCGCATCGCGCAACCGCTTTTCGTCGCAA 4066
972 laAlaGlnleuGlyAsnArgleuGlyAlaLeuArgTyrleuLys..... 986
4067 AAGCGATTCAGCATTCGAAACGTCATATCGCACCCCGCGGCTGCA 4116
987 .....ThrProGluThrAlaGluMetAlaGlyProGluAlaIleGlyAl 1000
4117 TTCACCGCTACCGCGCGGCGATTAAAGCAGATTATTCGCAACCGCG 4166
1000 aIleSerAlaIleValAlaIleGlyAspGluAlaIleAspAsn..ValAl 1016
4167 GCACACATTCGCTACCGCGCTTATTTGAGCCTGTCCTATACGATGCC 4215
1016 aTyrGlyIleThrAlaLysProPhe.....TyrThrAspAlaH 1029

```

```

4216 .....GTTCCGCAAGTCCGACGCG... 4239
1029 tsGlnSerLysGlyGlyLeuAlaGlyTyrLysAlaLysThrThrGly 1045
4239 ..... 4239
1046 ValValIleGlyLeuAspThrleuAlaAsnAspAsnleuMetIleGlyAl 1062
4240 .....GTCAATACCGCGCTATTTGGCGGAGATTTCGGCAAA 4276
1062 aAlaIleGlyIleThrLysThrAspIleLysHisGlnAspTyrLysLys. 1078
4277 CCGCAGTGGCGAATGGCGGCTTAACGCGCAATCAAGATTTCACGCTG 4326
1079 .....GlyAspLysThrAspAlaLysnGlyPheSerPhe 1089
4327 TCCTCCACGCTGCGCGCGCGCAAGGCGCGCAATTGGAACGCGACAG 4376
1090 SerleuTyrGlyAlaGlnGlnleuValLysAsnPhEAlaGlnGlyse 1106
4377 CGCG 4380
1106 rAla 1107

seq_name: SwissProt_40:PMPB_CHLMU
seq_documentation_block:
ID PMPB_CHLMU STANDARD; PRT; 1672 AA.
AC Q9PUY2;
DT 16-OCT-2001 (Rel. 40, Created)
DR 16-OCT-2001 (Rel. 40, Last sequence update)
DR 16-OCT-2001 (Rel. 40, Last annotation update)
DE Probable outer membrane protein pmpB precursor (Polymorphic membrane
DE protein B).
GN PMPB OR TC0694.
OS Chlamydia muridarum.
OX Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
NCBI_Taxid=83560;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MOPN / N199;
RX MEDLINE=20150235; PubMed=10684935;
RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidberg J.F.,
RA White O., Hickey E.K., Peterson J., Utterback T., Berry K., Baas S.,
RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
RA Gwin M., Nelson W., Deboy R., Kolonay J., McClarty G., Salzberg S.L.,
RA Eisen J., Fraser C.M.;
RT "Genome sequences of Chlamydia trachomatis MOPN and Chlamydia
RT pneumoniae AR39."
RL Nucleic Acids Res. 28:1397-1406(2000).
CC -!- SUBCELLULAR LOCATION: CELL WALL SURFACE (ELEMENTARY BODIES)
CC (POTENTIAL).
CC -!- SIMILARITY: BELONGS TO THE PMP OUTER MEMBRANE PROTEIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: AE002338; AAF39510.1; -.
CC TIGR: TC0694;
DR InterPro: IPR003357; OMP.
DR Pfam: PF02385; OMP; 1.
KW Outer membrane; Signal; Multigene family; Complete proteome.
FT SIGNAL 1 14
FT CHAIN 15 1672 POTENTIAL.
FT CHAIN 15 1672 PROBABLE OUTER MEMBRANE PROTEIN PMPB.
SQ SEQUENCE 1672 AA; 176295 MW; 4ABF190DA4DF8BD6 CRC64;

```


alignment_scores:

Quality: 203.50 Length: 1139
 Ratio: 0.398 Gaps: 55
 Percent Similarity: 44.864 Percent Identity: 20.720

alignment_block:

US-09-303-518D-653 x PMPB_CHLMU ..

Align seg 1/1 to: PMPB_CHLMU from: 1 to: 1672

```

706 AATGATGACAGTGTGGCAGACAGTCACTTAGTAGCGAAAAATTAACA 755
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
246 AsnGlyAsnGlyAla.....AspSerGlnThrProSerHl 257
756 TAACCATATGCTTTTACCACAGAGCGCTCATTTGGGAGAGCGCT 805
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
257 slvspro.....GlySerGlyAlaIleTyrAlaThrGlyAla 270
806 CACCAATGTTTATCTATGATCCCAAAAGCAAAAGTGAATTAATG 855
    ::::|:|||||:|||||:|||||:|||||:|||||:|||||:|||||
270 sp...LeuThrIleSerAspSerGlnGluIleValPheSerValAsnLys 285
856 GTATTGCAACAGCAACCCCTTATAGAAAAAGCAAT...GGCTTCCA 902
    ::::|:|||||:|||||:|||||:|||||:|||||:|||||:|||||
286 AlaSerLysAspGlyGlyAlaIlePheAlaGluLysAsnValSerPheG 302
903 G.....CTAGTCGTAAGATGGTCTCATGATGA..... 933
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
302 uAsnIleThrThrLeuLysValGlnAsnGlnAlaGluGluLysGlyG 319
934 ....ATCTTGGCTGAGATPACCATTCAGTATTCAGAACCATCA 978
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
319 LysIleTyrAlaSerGlyAspLeuSerIle.....Gln 330
979 AATGGAATATAC...TTTTTAACGACATATAAT.....CG 1013
    ::::|:|||||:|||||:|||||:|||||:|||||:|||||:|||||
331 SerSerLysGlnSerLeuPheAsnSerAsnThrSerLysGlnGlyGly 347
1014 CGCA.....GGAATAATCGATGCAAAACAT...AAACACTAT 1048
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
347 yAlaLeuTyrIleGluGlyAsnValAspPheLysAspLeuGluGluIle 364
1049 CTCCTACTATAGATTAAAAACAGCAACCGTTCAATGTTTATGTTCT 1098
    ::::|:|||||:|||||:|||||:|||||:|||||:|||||:|||||
364 rGileLysTyrAsn...LysSerGlyThrPheGluThrLysLysValThr 379
1099 TTATCC..... 1104
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
380 LeuSerLeuProGluAlaGlnThrAsnLysSerSerValThrAlaLase 396
1105 .....GACACAGCAAGACAGACCTGTTATCATG 1132
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
396 rGlnSerGlyProAsnThrThrProThrProThrProValThrAlaL 413
1133 CTGACAGTGGGTCAACAGTTATTCACCCAGCACTGAATAAGAAAAT 1182
    ::|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
413 ysgGlyGlyGly.....LeuTyrThrGluLysAsn 422
1183 ATTTCCTTATTTACAAAGAAAAGTGAATGATGATCTTACAGCAAC.. 1230
    ::|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
423 LeuSer...IleSerAsnIleThrGlyIleIleGluIleThrAsnAsnLys 438
1231 .ATCAACCAAGCGCGCGCGCTTTATATTGAGGCTAATTTTACGGTCT 1279
    ::|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
438 sAlaThrAspValGlyGlyGlyAlaIleValLysGlyThrLeuThrCysL 455
1280 CG.....CCTAAACAACGAAACGTCGCAAGCG 1308
    ::|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
455 yAspSerHisArgLeuGlnPheGlnLysAsnSerSerGluLysLysGly 471
1309 GCGGCGCTTCATATGATGATGCGAGTACCTGTTGTTGAAAGTAAACGG 1358
    ::|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
472 GlyGlyLeuTyrThrGluAsp.....ThrIleThr..... 481

```

```

1359 CGTGGCAAAACGACCCCTGTCCAAATGCGCAAGACAGCTGTGTTTC 1408
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
482 .....LeuSerAsnLeuThrGlyLysThrLeuPheGln 493
1409 AA.....GCCAAGGGGAAAAACCAAGCTCGGTACAGCTGGCGACGCT 1452
    ::|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
493 LuAsnThrAlaLysGluGluGlyGlyLeuTyrIleGlnLysAspAsp 509
1453 AAAGTCATC.....TTAGATCAG..... 1470
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
510 LysThrLeuThrMetThrGlyLeuAspSerPheCysLeuIleAspAsnTh 526
1471 .CAGCGGACGATCAAGCAAAACAAACCTTAGTGAATTCGCG.. 1515
    ::|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
526 rSerAlaThrHisGlyGlyAlaTyrValThrLysGluIleSerGlnT 543
1516 .....TTGGTCAGCGGCGAGGGAGCGTGCACACTGAATGCC 1551
    ::|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
543 hTyrThrSerAspValGluGluPheProGlyIleThrProValHisGly 559
1552 GATAATCAGTTCAACCCGACAAACTATTCGCTTTCGCGCGCA.. 1599
    ::|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
560 GluThrIleIleSerGlyAsnLysAlaThrGlyLysSerGlyGlyAla 576
1600 .....CGTTGGATTGAACGGGCGATCGCTTTCGT 1630
    ::|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
576 lCysThrLysHisLeuValLeuSerAsnLeuGlnThrIleSerIleSerG 593
1631 TCACCCGATTCAAATAATCCGATGAGGGGCGATGATGTCACACCAAT 1680
    ::|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
593 LuAsnPheAlaSerGluAsnGlyGlyAlaCysThrCysProAspAsn 609
1681 CACAGCAAGAATCCACCGCTTACCATTAACAGCAAT..... 1716
    ::|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
610 PheProAlaProThrAlaSerThrProSerThrAsnGlnThrAlaLapr 626
1717 .....AAAGAT.....ATTACACCAACCGCA 1738
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
626 OlyAspAspLysAspPheLeuIleAspTyrValValSerThrThrIle 643
1739 ATACACCAACTGTGATAGCAAAAGAAATGCGCTCAACGGTGGTTT 1788
    ::|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
643 sPlyAsnLysAlaThrLysLysGlyAla.....GlyValTyr 655
1789 GCGCAGAAAGATGCAACCAACGACGCGGCTCATCTGAAT..... 1833
    ::|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
656 AlaLysLysAlaLysLeuSerArgIleAspGluLeuAsnIleSerAsp 672
1834 .....TACCAACCGGAGAG 1849
    ::|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
672 nAlaAlaGlnGluThrGlyGlyPheCysCysThrGlnSerLeuGluL 689
1850 CGGATCCGACTTACTCTTTCGCGGCAACAAATTTAAAGGC..... 1893
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
689 euAspThrIleAlaSerLeuSerValThrLysAsnIleAlaGlyGlu 705
1894 .....AATATACGCAACCAACGCA 1916
    ::|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
706 GlyGlyLysLeuHisAlaLysThrLeuAsnIleSerAsnLeuLysSer 722
1917 ACTGTTTTTCAGC..... 1929
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
722 yLeuSerPheSerAsnAsnThrAlaAsnSerSerThrGlyValAlaLar 739
1929 ..... 1929
739 hThrAlaThrThrSerGlnSerProThrValSerSerPheLeuProArg 755
1930 .....GCGACAGCGACACCGGACGCGCTA 1952
    ::|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
756 AlaThrAlaLysSerSerProAlaProAlaGlnThrThrProThrTyrAl 772

```



```

FT CHAIN 22 1723 PROBABLE OUTER MEMBRANE PROTEIN PMF20
FT DOMAIN 94 97 POLY-SER.
FT DOMAIN 140 144 POLY-SER.
FT DOMAIN 147 160 POLY-SER.
FT CONFLICT 1134 1134 I -> V (IN REF. 1).
SQ SEQUENCE 1723 AA, 179594 MW, 4507D558A0E3EF01 CRC64;

alignment_scores:
    Quality: 203.50 Length: 985
    Ratio: 0.461 Gaps: 49
Percent Similarity: 44.772 Percent Identity: 21.218

alignment block:
US-09-303-518b-653 x PM20.CHILPN ..

Align seg 1/1 to: PM20.CHILPN from: 1 to: 1723

```

[illegible]

```

1256 ATTTGAGGTAATTTTACGGCTCG..... 1281
1259 YTCYSGLYASPAAPValThrLeuThrAsnLeuThrGlyLysThrLeuPhe 645
1282 CCTAAMAAACAAGAACGTCGACGAGCGGGCGCTCATATCAGTGATGG 1331
646 GlnGlnAsnSerSerGlnLysHisGlyGlyLysLeuSerLeuAlaSerGln 662
1332 CAGTACCCCTTCTTGGAAGTAAGCGGGTGGCAACACCGCCGCTGCA 1381
662 LysSerLeuThr.....MetHis 669
1382 AAATGGCAAGACGACGCTGCTGTTCAAGCAAGGGGAAACCAAGGC 1431
669 erLeuGlnSerPheCysLeuAsnAlaAsnThrAlaLysGlnAsnGlyLys 685
1432 TCGGTACAGCTGGCGGCGGTAAGTCACTTA..... 1464
686 GlyAlaAsnValProGlnAsnIleValLeuThrPheThrTyThrProTh 702
1465 ...GATACGACGCGGACGATCA.....GGCAAAACCAAG 1498
702 rProAsnGlnProAlaProValGlnGlnProValTyGlyGlnAlaLeu 719
1499 CCTTAGTGAATCGCTTGGTCACGCGGAGGCGGCGTCACTGAT 1548
719 alThrGlyAsnThrAlaThrLysSerGlyGlyLysLysAsn 735
1549 GCCGATATCACTCAACCCCGACAAACTCTATTTCGGCTTGGCGGG 1598
736 Ala.....AlaPheSer..... 739
1599 ACCTTTGATTTGAACGGGATTCGCTTCCACCGCATTCAAATA 1648
740 .....AsnLeuSerSerValThrPheAsp.....GlnAsn 750
1649 CCGATGAA.....GGGCGATGATTTGCACCAACATCAAGACAA 1689
750 hSerSerGlnAsnGlyAlaLeuLeuThrGlnLysAlaAlaAspLys 766
1690 GAA.....TCCACCGTACCATTCAGGCAAGAA 1718
767 ThrAspCysSerPheThrTyThrIleThrAsnValAsnIleThrAsnAsn.. 782
1719 AGATATTACTACACCGCATTAACAACAATCTG..... 1752
783 .....ThrAlaThrGlyAsnGlyGlyLysIleAlaGlyLysAlaAsn 797
1753 .....GATAGCAAAATA 1764
797 AspPheAspArgIleAspAsnLeuThrValGlnSerAsnGlnAlaLysLys 813
1765 .....GAAATTCGCTACAA 1778
814 GlyGlyGlyValTyLeuGlnAspAlaLeuIleLeuGlnLysValIleTh 830
1779 CGCTTGCTTGGCGAAGAAATGCAACCAAAAGAAAGCGGCGCTC... 1824
830 rGlySerValSerGlnAsnThrAlaThrGlnSerGlyGlyLysIleTyA 847
1825 .....AATCTGAATTACCAACCGGAAGAGCGGATCGCACTTACG... 1866
847 lalysAspIleGlnLeuGlnAlaLeuProGlySerPheThrIleThrAsp 863
1867 .....CTTTCGGCGGACAAATTTAAAGCGCAATAT 1898
864 AsnLysValGlnThrSerLeuThrThrSerThrAsnLeuTyGly.... 878
1899 CACGCAAAACAAACGCAAACTGTTTTCAGCGGCGACACCGACACCGCAG 1948
879 .....GlyGlyLysLeuSerSerGlyAlaValThrLeu.... 889

```

```

1949 CCTACATCATTTAGAACGCGGTGTCAAAAATGAGATATCCACAA 1998
890 .....ThrAsnIleSerGlyThrPheGlyIleThrGly 900
1999 GGAGAAATCGTGTGGACAACGATTCGATCGACCCGACATTTAAAGCGA 2048
901 AsnSerValIle.....AsnThrAlaThrSerGln 910
2049 AAACCTTCATATTCAGGCGGA.....CAGCGGTGTTCCTCCGCAATG 2092
910 nasPalaAspIleGlnGlyGlyLysIleTyAlaThrThrSerLeuSerI 927
2093 TTGCCAAATGGAAGCGATGTCATTTAAGCAATACCCGCCAGAGAT 2142
927 leaSnGlnCysAsnThrProIleLeuPheSerAsnAsnSerAlaAla.. 942
2143 TTGGGTGTCCACCGCATCAACACACAACTGTACAGTTCCGACTG 2192
942 ..... 942
2193 GACGGGTGTGCAAGTTGTACGAAAAAACCATTCACGACGATTAAGTG. 2241
943 .....ThrLysThrSerThrThrLysGlnIleA 953
2242 .....ATTGCTTCATTCAGCAAGACGACATTCAGAGGCAT... 2277
953 laGlyAlaIlePheSerAlaIleValThr...IleGlnAsnAsnSer 968
2278 .....GTCAAGCTTGGCGATCAGCTCATTTAATCTCAACAGACT 2318
969 GlnProIleIlePheLeuAsnAsnSerAlaLysSerGlnAlaThrAl 985
2319 TGCCACACTCAACGGCAATCTTGTGACGGCGA.....GACAGCG 2359
985 alAlaThrAlaGlyAsnLysAspSerCysGlyAlaIleAlaAlaAsn 1002
2360 ACTATACGTTTACCGGCAC.....GCCACCAAAAGCGCACTCAGC 2403
1002 erValThrLeuThrAsnAsnProGlnIleThrPheLysGlyAsnTyAla 1018
2404 CTCGTGGCAATGCCAAGACATTTAATCAAGCCACATTAACGGCAA 2453
1019 GlnThrGlyAlaIleGlyCysIleAsp.....LeuThrAsnGlySe 1033
2454 C.....ACATCGGCTTCGACATGCTCTCAATTAATCTAAGCA 2491
1033 rProProArgLysValSerIleAlaAspAsnGlySerValLeuPheGlnA 1050
2492 ACAAGCGCGTACAAAC..... 2508
1050 spAsnSerAlaLeuAsnArgLysAlaIleTyGlyGlnIleThrIleAsp 1066
2509 .....GGCAGCTGACGCTTCCGACAAAGCTTAAGCAACGT 2546
1067 IleSerArgThrGlyAlaThrPheIleGlyAsnSerSerLysAspGln 1083
2547 AAGCATTCGCCCACTCAAGCGCATGCTCCCTAAGCCGATTAAGCACTAT 2596
1083 ySerAlaIleCysCysSerThrAlaLeuThrLeuAlaProAsnSerGlnL 1100
2597 TCCATTTTGAAGAACCGGCTTACCGGAAATCAAGCGGCGGCAAGAT 2646
1100 euIlePheGlnAsnAsnLysValThrGlnThrAlaThrThrThyAla 1116
2647 ACGCATTTACACTTA.....AAAGCAGCGA 2672
1117 SerIleAsnAsnLeuGlyAlaAlaIleTyGlyAsnAsnGlnThrSerAs 1133
2673 ATGACAGCTGCGCGTCCGACGGAATTAAGCAATTAACCTTGACAAAG 2722
1133 pIleThrIleSerLeuSerAlaGlnAsnGlySerIlePhePheLysAsnA 1150
2723 CCACCATTAACACTCAATTCGCTATGACACAGATGCGGACGCGCCAA 2772

```

```

1150 snleucysttrhlatrhnsllystfycysserllealaglasyanvallys 1166
2773 ..ACCGGACAGTGGCGAGATGGCGCGCGCGCGTTCGGCGCGTCCCT 2819
1167 PhehtrAlaIleGlAlaSerAlaGlyLysAlaIleSerPheTyrrspAl 1183
2820 ATATATCGGTATAGCGCGGACATTCGGCGAGAAATCCGTTTCAACACGCTGA 2869
1183 aValasNval.....SerThrlysgluThrAsnAlaGluLeuL 1197
2870 CGGTAAACGGCAATTCAGCGGTCAGGAAATCCGCTTTATGTGCGGA 2919
1197 yslenuasngluylsAlaThrSerThrglyThrleuPheSerGlyGlu 1213
2920 CTCTTCGGGTACCGCAGC.....GG 2939
1214 LeuHISgluAsnlySerTyrrleProGluLysValThrPheAlaHISgl 1230
2940 CAATATGAACTGGCGGAAATTCGGA..... 2967
1230 yasnleuIleleuGlyLysasnAlaGluSerValValSerPheThrg 1247
2968 .....GGCACTTACAC..... 2979
1247 lnsrProGlyThrThrIleThrmecGlyProGlySerValleuSerAsn 1263
2980 .....TTGGCTGTCAACAAATACGCGACAGCAAC 3008
1264 HIsSerlysgluAlaGlyIleAlaIleAsnAsr..... 1275
3009 CGTAATCTTCGAGCAATTCAGCTAGTGAAGAAAGAACACACACCGC 3058
1276 .ValIleIleaspheSerGluIleValProThrlyAspAsnAlaThrV 1292
3059 TGTGCC 3063
1292 :AATA 1293

seq_name: SwissProt_40:WAPA_BACSU
seq_documentation_block:
ID WAPA_BACSU STANDARD; PRT; 2334 AA.
AC Q07833.1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Wall-associated protein precursor.
GN WAPA OR N17G.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RC STRAIN=168 / BGSC1A1;
RX MEDLINE=95219088; PubMed=7704263;
RA Yoshida K.-I., Sano H., Seki S., Oda M., Fujimura M., Fujita Y.,
RT "Cloning and sequencing of a 29 kb region of the Bacillus subtilis
RT genome containing the hut and wapa loci."
RL Microbiology 141:337-343(1995).
RP SEQUENCE FROM N.A.

RC STRAIN=168 / BGSC1A1;
RX MEDLINE=97124196; PubMed=8969509;
RA Yoshida K.-I., Shindo K., Sano H., Seki S., Fujimura M., Yanai N.,
RA Mawa Y., Fujita Y.,
RT "Sequencing of a 65 kb region of the Bacillus subtilis genome
RT containing the lic and cel loci, and creation of a 177 kb contig
RT covering the gnt-sacry region."
RL Microbiology 142:3113-3123(1996).
CC -1- FUNCTION: STILL UNKNOWN. NOT INVOLVED IN CELL MEMBRANE METABOLISM,
CC MOTILITY, SECRETION OR DIFFERENTIATION.
CC -1- SUBCELLULAR LOCATION: MEMBRANE-ASSOCIATED. MAY BE RELEASED
CC INTO THE MEDIUM.
CC -1- DOMAIN: HAS TWO LIGAND-BINDING DOMAINS: THE N-TERMINUS, HAS THREE
CC 101 AA REPEATS WHICH ARE RESPONSIBLE FOR CELL WALL BINDING; THE
CC C-TERMINUS CONSISTS OF TWO BLOCKS OF RESIDUES WITH A CONSERVED
CC MOTIF REPEATED 31 TIMES.
CC -1- SIMILARITY: THE REPEATED MOTIF IN THE C-TERMINUS HAS SOME
CC SIMILARITY TO THE REPEAT IN E. COLI RBS GROUP OF PROTEINS (RBSA-D).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L05634; AAA22883.1; -
DR EMBL; D31856; BAA0656.1; -
DR EMBL; D29985; BAA06260.1; -
DR EMBL; D83026; BAA11683.1; -
DR EMBL; Z99124; CAB15959.1; -
DR PIR; S32920; S32920.
DR Subtilisin; BG10797; wapa.
DR InterPro; IPR003305; CBD_6.
DR Pfam; PF02018; CBD_6; 1.
KW Cell wall; Repeat; Signal; Complete proteome.
FT SIGNAL 29 28
FT CHAIN 1 2334
FT DOMAIN 504 869
FT REPEAT 504 605
FT REPEAT 636 736
FT REPEAT 769 869
FT DOMAIN 1021 2139
FT REPEAT 1021 1040
FT REPEAT 1042 1061
FT REPEAT 1063 1082
FT REPEAT 1083 1102
FT REPEAT 1109 1128
FT REPEAT 1129 1148
FT REPEAT 1150 1169
FT REPEAT 1174 1193
FT REPEAT 1199 1218
FT REPEAT 1219 1238
FT REPEAT 1246 1265
FT REPEAT 1267 1286
FT REPEAT 1287 1306
FT REPEAT 1309 1327
FT REPEAT 1329 1348
FT REPEAT 1349 1369
FT REPEAT 1370 1390
FT REPEAT 1391 1410
FT REPEAT 1411 1430
FT REPEAT 1431 1450
FT REPEAT 1451 1470
FT REPEAT 1471 1490
FT REPEAT 1491 1510
FT REPEAT 1511 1530
FT REPEAT 1531 1550
FT REPEAT 1551 1570
FT REPEAT 1571 1590
FT REPEAT 1591 1610
FT REPEAT 1611 1630
FT REPEAT 1631 1650
FT REPEAT 1651 1670
FT REPEAT 1671 1690
FT REPEAT 1691 1710
FT REPEAT 1711 1730
FT REPEAT 1731 1751
FT REPEAT 1752 1772
FT REPEAT 1773 1792
FT REPEAT 1793 1814
FT REPEAT 1815 1839
FT REPEAT 1840 1859
FT REPEAT 1861 1880
FT REPEAT 1881 1906
FT REPEAT 1907 1927
FT REPEAT 1928 1948
FT REPEAT 1949 1968
FT REPEAT 1969 1982
FT REPEAT 1983 2002
FT REPEAT 2003 2027
FT REPEAT 2028 2047
FT REPEAT 2048 2070
FT REPEAT 2051 2070

```


693 laLeuHisLeuLysAspGlySerGlyAlaGluLeuProIleAsnProGly 709
1294 GAACCTGGCAAGCGCGGCGC..... 1314
710 ProThrTyrLysAsnAlaGlyLysPglYAlaLysArgAsnTyrSerPh 726
1315GTTCATATCAGTGAATGGCAGTACCGTTACTTGGAAAGTAA 1354
726 eLysIleIleAlaTyrAsnLysAspGlyLualle.....Alas 740
1355 ACGGCGTGGCAACGACCGCTGTCCAAATATGGCAA..... 1392
740 eProAlaIleThrProAlaLeuProAspIleAlaArgProLysAsnVal 756
1393 ...GGACCGCTGTGGTTCAGCCAAAGGGGAAACCAAGGC..... 1431
757 ThrGlyTyrLeuTyrThrAsnThrLysSerSerGlnThrGlyTyrValAs 773
1432TCGCTCAGCGTGG 1444
773 nLeuIleTpgLysValGlnAsnAlaLysGlyTyrLysValAsnIleT 790
1445 GCGAGCGTAAAGTCATCTTAGTCAGCAGCGGACGATCAGCAAAAAA 1494
790 yrasnGlyLys.....GluTyrGlnSerPheAspValGlyAspAla 803
1495 CAAGCCTTATAGT.....GAAT 1511
804 AspHisTprThrThrGlnAsnLysAsnIleTprThrSerGlnIle 820
1512 CGGCTTGTCAGCGGAGGAGCGGTGCACATGATCCGATTAATCAG. 1560
820 e.....LysAlaGlySerTyrLysLeuHisThrAspGlyLysG 833
1561TTCAACCCCGCAAACTCTAT..... 1581
833 LysGlyLeuLeuAlaLeuAspProSerProValTyrAsnAsnAlaAsnGly 849
1582TTGCGCTTTCGCGGCGGACGTTTGATTT 1610
850 AsnTyrLysGlyLysLysAsnTyrSerPheThrLeuValAlaTyrAspAl 866
1611 CAAGCGGCAATTCGCTT.....TCGTTCCACCGCA 1639
866 aasnGlyLutThrIleProThrAlaProPheAsnProThrPheHis... 881
1640 TTCAAATAATACGATGAGGGGCGATGATTCACACACATCAGACAAA 1689
882GluGlyAlaGluPheLeuGly...ThrGluLeuTyr 892
1690 GAATCCACCGTTACCATT.....ACAGGCAA 1715
893 TrpSerIleIleAspIleProSerGlyLysLeuAsnGlyAlaThrGlyLys 909
1716 T.....AAGATATTACTACAAACGCGCAATAC..... 1743
909 nValIleValAsnGluAspLeuSerIleAspGlyArgGlyProGlyL 926
1744AACACTGTGATGCAAAAAAAGAAATGCC 1773
926 euGlyLeuSerArgThrTyrAsnSerLeuAspSerSerAspHisLeuPhe 942
1774 TACAACGGTGTGGTGGC..... 1791
943 GlyGlnGlyTprTyrAlaAspAlaGluThrSerValIleSerThrAspGl 959
1792GAGAAAGATGCAAC.....A 1807
959 yGlyAlaMetTyrIleAspGluAspAlaThrThrHisArgPheThrLysL 976
1808 AAACGAAAGCGGCGCTCAATCTGAATTAATACCAACG..... 1842
976 ysaLAspGly.....ThrTyrGlnProProThrGlyValTyr 988
1843GAAGAAGCATCGCACTTACTGCTTCCGGCGAAC 1880
989 LeuGluLeuThrGluThrAlaAspGln...PheIleLeuLysThrLysAs 1004
1881 AAATTTAAACGGCAATATCACGCAAAACAAAGCGGAACTGTTTCAGCG 1930
1004 pGlnThrAsnAlaTyrPheAsnLysLysGlyGlyLysLeu.GlnLysVal 1020
1931 GCAGACCGACACCGCACCGCTACATC..... 1957
1021 ValAspGlyHisAsnAsnAlaThrValTyrThrTyrAsnAspLysAsnGl 1037
1958ATTAGAGAGGGGTGTCAAAAATGGAAGTATCCACA 1997
1037 nLeuThrAlaIleThrAspAlaSerGlyArgLysLeuThrPheThrTyrA 1054
1998 AGGAGAAATCGTG..... 2010
1054 spGluAsnGlyHisValThrSerIleThrGlyProLysAsnLysVal 1070
2011TGGACAACATTTGGATC.....GACCG 2033
1071 ThrTyrSerTyrGluAsnAspLeuLeuLysValThrAspThrAspGl 1087
2034 CACATTTAAAGCGGAAACCTTCATTCAGGGCGGCAACGCGTGT 2083
1087 yThrValThrSerTyrAspTyrAspSerGlnGly..... 1098
2084 CCCCAATGTTCGCAAGTGAAGGCGATGGCATTAACCAATCAGCGC 2133
1099 ..ArgLeuValLysGlnTyrSerAlaAsn.....SerThrGluAla 1111
2134 CAACGACTTTTCGGTGTGCGACCGCATCAAGCGACAACTGTACACG 2183
1112 LysProValPheThrGlnTyrGlnTyrSerGlyHisArgLeu..... 1125
2184 TTCGACTGAGCGGCTGTACAAAGTTGTACGAAAAACCATAC.... 2229
1126GluLysAlaIleAsnAlaL 1132
2230GACGATAAAGTATGCTTCA 2250
1132 yslLysGluThrTyrValTyrSerTyrAspAlaAspLysLysThrLeu 1148
2251 TTGACGACAGCCGACATCAGA.....GG 2273
1149 MetThrGlnProAsnGlyArgLysValGlnTyrGlyTyrAsnGluAlaGl 1165
2274 CAAT...GTCAGCGTTCGATCAGCGCTCAT...TTAAATCTCAGAGAC 2317
1165 yAsnProIleGlnValIleAspAspAlaGluGlyLeuLysIleThrThra 1182
2318 TTGCGACACTCAACGCGCAATCTTAGTCAGCGGAGACACGCACTATAGC 2367
1182 snThrLysTyrGluGlyAsn..... 1188
2368 GTTACGGGCAACGCCCAACCGCAACCTCAGCTGTGGGCAATGC 2417
1189AsnValValGluAspValAspProAsnAspAlaGlyThrGln 1202
2418 CCAAGCAACATTTAAT...CAAGCCACATTAACGCAACACATCGGCTT 2464
1202 yLysAlaIleThrGlnSerTyrGlnTyrAspLysAspGlyAsnValThrSerV 1219
2465 CGGACAATGCTTCATTTAATCTAAGCAACACGCGGTACAAAAGGCACT 2514
1219 alLysAspAlaTyrGlyThrGlnTyrThrTyrGlnTyrAsnLysAsnAsp 1235
2515 CTGACGCTTCGCAACGCTAAGCAACGTAAGCATTCGCGCACTCA 2564
1236 ValThrLysMetLysAspThrGlnGlyAsnValThrAspIleAlaTyrAs 1252

```

2565 CGGC.....ATGTCCTCCAGCGATAG.....2589
      ||| ||| ||| |||
1252 pclyleuaspalaValSerGluThrAspGlnSerGlySerSerSera 1269
2590 ..GCAGTATCCATTTTGAAGACCGCGTTTACCGGAAATAAGCGCG 2637
      ||| ||| ||| |||
1269 laalavalYr.....AspLysTYGlyAsnGlnIleGlnSer 1281
2638 GCGAGGATACCGCATACAC.....TTAAAGAC.....2667
      :||| ||| ||| |||
1282 SerLysAspLeuSerAlaSerThrAsnIleLeuLysAspGlySerPheG 1298
2668 .....AGCAATGAGCGCTCCGCGGCGACGGAATTAGGCAT 2707
      ||| ||| ||| |||
1298 ualaglInlysserGlyTrpAsnLeuThrAla.....1308
2708 TAAACCTGACACGCCACATTTACATCAATTCGCGCTATCGACAGAT 2757
      :||| ||| ||| |||
1309 ..SerLysAspArgArgLysIleSerValIleAla.....Asp 1320
2758 GCGGAGGCGCGCAACCGCGAGTGCAGATGCGCGCGCGCGCTTC 2807
      :||| ||| ||| |||
1321 LysSerGlyValLeuSerGlySerLysAla.....1330
2808 GCGCGCTCCCTATATCCGTTACCGCGCACTTCGCGAGAACCCGTT 2857
      :||| ||| ||| |||
1331 ..LeuGluValLeuSerGlnSerThrSerAlaGlyThrAspHisGly 1346
2858 TCACACACGCTGACGTAACGCAATTCAGCGTCACGAGAACNTCCGC 2907
      :||| ||| ||| |||
1346 YrSerSerAlaThrGlnThrValGluLeuGluProAsnThrThrThr 1362
2908 TTTATGTGCGACCTTCGCGTACCGCGCAAGCAATTCGAG.....CT 2951
      :||| ||| ||| |||
1363 Leu.....SerGlyLysIleLysThrAspLe 1371
2952 GCGGGAAGTTCCGAGGCACTTACACCTTGCGTTCACAACTACCGGCA 3001
      :||| ||| ||| |||
1371 ualalys...SerArgAlaTyRPhaAsnIleAspLeuArgAspLysAsp 1387
3002 AGCAACCCCTAGTCTC.....GAGCAATTACGCTAGTGAAGGAA 3045
      :||| ||| ||| |||
1387 InLysArgIleGlnThrPheHisAsnGluYrSerAlaLeuAlaGlyLys 1403
3046 GACACACACCGCTGTCGAAATCTTAATTCAC.....3081
      ||| ||| ||| |||
1404 ..AsnAspTrpThrLysArgGlnIleThrPheThrThrProAlaAsnAl 1419
3082 .....CTGCAAAACGACACGTCGAT.....3102
1419 agLysAlaValAlaValTyMetGluValAspHisLysAspLysAspGlyL 1436
3103 ..GCGGCGCATGCGGTTATTCAGCTTATCCGCAAGAGCGGAGTTC... 3147
      :||| ||| ||| |||
1436 ysgLysAlaThrPheAspGluValGlnLeuGlnLysGlyLysValSer 1452
3148 CCGCTGCATATCCGCTCAAGACAGCTTTCGCAAACTC.....3192
      :||| ||| ||| |||
1453 SerSerTyAsnProValGlnAsnSerSerPheThrSerAlaThrGluAs 1469
3193 .....GCGAGCGCGGAGAAACAGAGCGCGCTTGACGCGCA 3229
      ||| ||| ||| |||
1469 nTrpAsnValSerGlyAlaSerValAspSerGluGlnTyPheAsnAsp 1486
3230 AACAGGCACAACTGCGCGCAAAACAAACGCGGAAAGACAAAGCGGCA 3279
      :||| ||| ||| |||
1486 spValSerLeuLysAlaAlaArgThrSerAlaSerGlnAlaLysSerVal 1502
3280 ACCCTTGACGCGCTGATTCGCGCGCGCAATTCACGCAAGAAAGCGCA 3329
      :||| ||| ||| |||
1503 ThrLysGlnThrValVal...LeuGlyGlnSerAlaAsnAspLysProVa 1518
3330 AAGTGTGCGGAACCGCGCGGCGAGCGCGGGAATAATTCGCGCATTA 3379

```

```

1518 lTyLeuThr.....LeuThrGlyMetSerLysAlaSerSerV 1531
      :||| ||| ||| |||
1531 alLysPheThrAspGluLysAspTYrSerLeuGlnIleAsnValThrTYr 1547
3427 GCCTTGCGAAGCGCGGAGCGGAACCGCGCGCTTCACCGCCTT 3476
      ||| ||| ||| |||
1548 Ala.....AspGlySerThrGlyIleTyAsnAlaLysPhe 1559
3477 CCCCCCGCGC.....CGCGCGCGCGCGCGGATTTGCGCGAAC 3514
      :||| ||| ||| |||
1559 eProSerGlyThrGlnGluTrpAsnArgAlaAlaValIleProLysTr 1576
3515 CGAGCGCCCAACCGCAACCCCAACCGCAGCGGACCTGATACCGCTAT 3564
      :||| ||| ||| |||
1576 hLysPro.....1578
3565 GCCAATAGCGGTTTGATGATTTTCGCGCACGCTCACAGCGTTTCCG 3614
      :||| ||| ||| |||
1579 .....IleAsnLysVal.....1582
3615 CGTACAGACGAATTCGACCGCGCTTTGCCGAAGACCGCGCAAGCGCG 3664
      :||| ||| ||| |||
1583 .....AspLysIleLeuPheGlnLysSerAlaThrGlyThrV 1596
3665 TTGGACACACGCGCATCCG.....3684
      :||| ||| ||| |||
1596 alTrpPheAspAspIleArgLeuIleGluGlySerLeuThrLysSer 1612
3685 .....GACACCAACACACTAC.....CGTTCGCAAGATTTCCGCGC 3719
      :||| ||| ||| |||
1613 ThrTyAspSerAsnGlyAsnTyValThrLysGluGlnAspGluLeuG 1629
3720 CTACCGGCAACAAACGACCTGCGCCAAATCGGTATGAGAAATAC.... 3765
      :||| ||| ||| |||
1629 YTYrAlaThrSerThrAspTyAspGluThrGlyLysLysThrSerGlu 1646
3765 .....3765
1646 hrAspAlaLysGlyGluLysThrThrTyThrTyAspGlnAlaAspGln 1662
3766 .....CTGCGACGCGGCGCGTCCGATCTGTTTCGCA 3800
      :||| ||| ||| |||
1663 LeuThrAsnMetThrLeuSerAsnGly...ThrSerIleuHisSerTy 1678
3801 CAACCGGACCGGAAAC 3816
      :||| ||| ||| |||
1678 rAspLysGluGlyAsn 1683
seq_name: SwissProt_40:VAC3_HELPY
seq_documentation_block:
ID VAC3_HELPY STRAND: PRT: 1310 AA.
AC Q48253;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Vacuolating cytotoxin precursor.
GN VACA.
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
OX NCBI_TaxId=210;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TX30A;
RX MEDLINE=9535366; PubMed=7629077;
RA Atherton J.C., Cao P., Peek R.M., Jr., Tummuru M.K., Blaser M.J.,
RA Cover T.L.;
RT "Mosaicism in vacuolating cytotoxin alleles of Helicobacter pylori.
Association of specific vacA types with cytotoxin production and

```


1783TGTTGGGAGAGAAAGTCAACCAAAAGCGGGGGCGTCAATC 1828
1784|||||.....
516 TAlAhIsPheLySAsnIleAsPAlAtHrLySerAsPAsnGIleuIsnt 533
1829 TGAATTACCAACCGGAAGAGGATCCACTTACTGCTTTCCGGGGGA 1878
533 hrser.....ThrLeuAsPheSerGIleu 541
1879 ACAAAAT.....TTAAACGGCAATATACGCAAAACAA 1910
542 ThrAsPlySeValAsnIleAsnLySerThrAlaAlAtHrAsnValAs 558
1911 CGGCAACAGCTTTT...CACGGGACAGCCGACACCGGACCTTACAAT 1956
558 nIleLySAsnPhAsPleAsPleLySgluLeuValAlThrArGIValGlns 575
1957 CATTTAGGAAGCGGGGTCAAAATGGAAGGTATCCACAAAGAGAAT 2006
575 erphe.....GlyGIInTyTrHrIlePheGIleu 584
2007 CGTGTGGACAACGATTTGGATGACCGGACATTTAAACGGGAAACCTCC 2056
585 AsnIleGIyAsPlyS...SerArGIleGIyAlValSerIleuGIInHrGI 600
2057 ATATTAGGGCGGACACAGCGGTGGTTCCCGCAATGTGCCAAAGTGAA 2106
600 yTySerProAlAtySerGIyGIleVal...ThrPheLySgluGIle 615
2107 GGGGATTTGGCATTTTAAAGCAATCAGCCCAAGAGATTTCCGTGCGCAC 2156
615 yslLySeruValIle.....619
2157 GCATCAAGCCACACAACTGTACACTGTGAGCTGACGGGTGTGACAA 2206
620AspGI 621
2207 GTTGTACCGAAAA...AACCATTTACCAGAGATTAAGTATTTGCTTCA 2250
621 uIleTyTrHrAlAsPAlAtProTyPrAsnTyPheAsPAlAtArGIAsnVal... 635
2251 TTGAGCAACCGGACATCGAGGCAATGTCAAGCCCTGGCCGATCAAGCTCA 2300
636ThrAsPlyAlGIleAsnLySarGIleuPheGIyAlArP 649
2301 TTTTAATCTC.....ACAGAGATT.....GCCACACTCA 2329
649 oGIyAsnIleAlGIyLySThrGIleuMetPheAsnAsnIleuHrIleuA 666
2330 AGCGCAATGTTTAGTGA....GGCGGACAGCGACATATAGCGTTAGC 2373
666 snSerAsnIleAsPleMetAsPlyGIyLySAsPlyAsPleuHrIleGIIn 682
2374 CGCAACGACCCCAAAACGGC.....AACCTGAGCCCTCGGGGCA 2414
683 GIyHrIsPheThrAsnAsnGIInLyThrMetAsnIleuPheValGIInAsP 699
2415 TGGCCACAGCAACATTTAAT.....CAAGCCATTTAAACGGGACAA 2455
699 yArGIValAlAtHrIleuAsnAlGIyHrIsGIInAlAsPleMetIlePheAsnA 716
2456 CATCGGCTTGGACAATGCTTCATT.....AAT 2484
716 snIleuValAsPserThrTrGIyPheTyLySProLeuIleLySleAsn 732
2485 CTAAAGCAACAGCGGTACAAAGCGGCTGACGCTTTCGCAACAGC 2534
733 AsnAlAlGIInAsnIleuThrLySAsnLySgluHrIleuValIleu...Va 746
2535 TAAGCA.....2541
746 LysAlAlArGIAsnIleAsPlyTrAsnIleuValGIyAlGIInGIyAlAsP 763
3542AACGTAAGCAATTCGCACTCAACGGCAATGTCTCCCTTACCGGAT 2586

```

1012 .....LysArgLeuGlnAla 1016
3415 GATAAAGACACCGCTTGGGAAACAGCGGAAACCGGCGCGC 3464
      ||||| ||||| ||||| |||||
1017 LeuLysAsp.....GlnArgPheAlaSerLeuGlnSerAl 1028
3465 TACACCGCGCTTC.....CCCGCGCGCGCGCGCGCGCG 3502
      ||||| ||||| ||||| |||||
1028 aAlaGlnValLeuTyrGlnPheAlaProLysTyrGlnLys.ProThrAsn 1044
3503 ATTGCGCGCAAC.....CGACGCCCGCAACCGCAACCG 3540
      ||||| ||||| ||||| |||||
1045 ValTrpAlaAsnAlaIleGlyGlyThrSerLeuAsnSngLysSerAl 1061
3541 CAGCGCGAC..... 3549
1061 aserLeuTyrGlyThrSerAlaGlyValaAlaPheAlaTyrLeuAsnGlyLuv 1078
3550 .....CTGATCAGCGCTTATGCAATAGCGGTTGAGTGAATTTCC 3591
      ||||| ||||| ||||| |||||
1078 alGlnAlaIleValaIleGlyGlyPheGlySerTyrGlyTyrSerSerPheSer 1094
3592 GCCACGCTCAACG..... 3606
      ||||| ||||| ||||| |||||
1095 AsnGlnAlaAsnSerLeuAsnSerGlyAlaAsnAsnThrAsnPhelGlyVa 1111
3607 .....GTTTCGCCCGTACAGAGCAATTTGAC.....C 3634
      ||||| ||||| ||||| |||||
1111 ITyrSerArgLeuPheAlaAsnGlnHisGlnPheAspPheGlnAlaGln 1128
3635 GCGTGTTCGCGAAGACCGCGCGCGCGCTTTGGACAGCGCG..... 3678
      ||||| ||||| ||||| |||||
1128 LysAlaLeuGlySerAspGlnSerSerLeuAsnPhelSerAlaLeuLeu 1144
3679 .....ATCGCGGACCAACAACTACCGTTGC..... 3705
      ||||| ||||| ||||| |||||
1145 GlnAspLeuAsnGlnSerTyrHisTyrLeuAlaTyrSerAlaThrThrAr 1161
3706 .....CAGATTTTCGCGCTTACCGCGCAACAAACGACCTGC 3742
      ||||| ||||| ||||| |||||
1161 gAlaSerTyrGlyTyrAspPheAlaPhePheArgAsnAlaLeuValLeuL 1178
3743 GCCAAATCGGTATGCAAAAAAAGCTGCGGAGCGCGCTCGGCATCTGT 3792
      ||||| ||||| ||||| |||||
1178 ysPro.....SerValGlyVal... 1183
3793 TTTTCGCAACAGCGGACCGGAAACCTTCGACAGCGCATCGGCACATC 3842
      ||||| ||||| ||||| |||||
1184 ...SerTyrAsnHisLeuGlySerThrAsnPhelLysSerAsnSerAsnG 1199
3843 GGCACGCGCTTGCACGCGTCCGTTTCGCGCATACGCGATCGGACAGT 3892
      ||||| ||||| ||||| |||||
1199 nValAlaLeuSerAsnGlySer..... 1206
3893 TCGACATCGGCATCAGCGCGCGCGGCTTTAGTACGCGGACGCTTTCA 3942
      ||||| ||||| ||||| |||||
1207 .....SerSerGlnHisLeuPhe 1212
3943 GAGCGCATCAGAGGCAAAATCCGCGCGCGCTGCTGCATTACGCGCATTC 3992
      ||||| ||||| ||||| |||||
1213 AsnAla.....AsnAlaAsnValG 1219
3993 GCGAAGATACCGCGAGTTTCGCGGATTCGCGCATCGAACCAGACATCG 4042
      ||||| ||||| ||||| |||||
1219 uAlaArgTyr.....TyrTyrG 1225
4043 GCGCAACGCGCTATTTGTCGCAAAACG 4071
      ||||| ||||| ||||| |||||
1225 LysAspThrSerTyrPheTyrMetAsnAla 1234
seq_name: SwissProt_40:PMPG_CHLTR
seq_documentation_block:

```

```

ID PMPG_CHLTR STANDARD; PRT: 1013 AA.
AC O84879;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Probable outer membrane protein pmpg precursor (Polymorphic membrane
DE protein G).
DE GN PMPG OR C7871.
OS Chlamydia trachomatis.
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=813;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=D/UM-3/CX;
RX MEDLINE=99000809; PubMed=9784136;
RA Stephens R.S., Kalman S., Lammel C.J., Fan J., Marathe R., Aravind L.,
RA Mitchell W.P., Olinger L., Tatusov R.L., Zhao Q., Koonin E.V.,
RA Davis R.W.;
RT "Genome sequence of an obligate intracellular pathogen of humans:
RT Chlamydia trachomatis."
RL Science 282:754-759(1998).
CC -! SUBCELLULAR LOCATION: CELL WALL SURFACE (ELEMENTARY BODIES)
CC (POTENTIAL).
CC -! SIMILARITY: BELONGS TO THE PMP OUTER MEMBRANE PROTEIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: AE001360; AAC68469.1; -
DR InterPro: IPR003368; DUF145.
DR InterPro: IPR003357; OMP.
DR Pfam: PF02415; DUF145; 1.
DR Pfam: PF02385; OMP; 1.
DR Outer membrane; signal; Multigene family; Complete proteome.
FT SIGNAL 1 27
FT CHAIN 28 1013 PROBABLE OUTER MEMBRANE PROTEIN PMPG.
FT SEQUENCE 1013 AA; 107366 MW; F0927743C0A651DD CRC64;

alignment_scores:
      Quality: 199.00      Length: 953
      Ratio: 0.431      Gaps: 49
      Percent Similarity: 48.478      Percent Identity: 21.301

alignment_block:
US-09-303-518D-653 x PMPG_CHLTR ..
Align seg 1/1 to: PMPG_CHLTR from: 1 to: 1013

1615 GGGCATTCGCTTTCGTTCCAGCATTCAGAAATACGATGAAGGGCGAT 1664
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
95 GlyHisSerLeuThrPheGlnLysnIleArgThrSerThrAsnGlyAlaAl 111
1665 GATTGTCAAACACATCAAGACAGAAAGATCCACCGTTACCATTTACAGCA 1714
      ||||| ||||| ||||| ||||| ||||| |||||
111 Aleu.....SerAspSerAlaAsnSerGlyLeuPheThrIleGlnGlyP 126
1715 ATTAAGATATTACTACACCGGCAATTAACACACTGTGATGCAAAAAA 1764
      ||||| ||||| ||||| ||||| ||||| |||||
126 heLysGlnLeuSerPheSerAsnGlyAsnSerLeuLeuAlaValLeuPro 142
1765 GAATGTCCTACACGCGTGTGCGGAGAAAGATGCAACCAAGCAAA 1814
      ||||| ||||| ||||| ||||| ||||| |||||
143 AlaAlaThrThrAsnAsn.....GlySerGlnThrProThrThrThSe 157
1815 CGGGCGGCTCAATCTGAATTACCAACGGAAGAGCGGATCGCACTTAC 1864
      ||||| ||||| ||||| ||||| ||||| |||||
157 r..... 157

```

```

1865 TGCTTCCGGCGAGACAATTTAAAGCGCAATATCACGCAAAACAGCGC 1914
158 .....ThrProSerAsnGlyThrIleTyrSerLysThrAsp 169
1915 AAACGTGTTTTCAAGCGCGAGACGACACCGGCGCTACACAATCATTTAG 1964
170 LeuLeuLeuLeuAsnGlnLysPheSerPhe...TyrSerAsnLeuVal 185
1965 AAGCGGTGTGTCAAAATGGAAGGTATCCCAAGAGAAATCGTGTGG 2014
185 IserGly.....A 188
2015 ACAACGATTGATCGACCGCATTTAAAGCGAAATCTTCATTTACG 2064
188 spGlyGlyAlaIleAsp.....AlaLysSerLeuThrValGln 200
2065 GGC.....GGACAAGCGGTGTTCCCGCAATGTTCCCAATGTGAAG 2108
201 GlyIleSerLysLeuLysValPheGlnGlnAsnThrAlaGlnAlaAspG 217
2109 CGAT.....TGGCATTTAAGCAATCAGCGCC 2134
217 yGlyAlaCysGlnValValThrSerPheSerAlaMetAlaAsnGlnAlaP 234
2135 AAGCAGTTTTC.....GGTGTGCAACCGCATCAAAAGCACAACATCTGT 2178
234 rolleAlaPheIleAlaAsnValAlaGlyValArgLysGlyIleAla 250
2179 ACAAGTTGCGACTGACGGGTGACAGAGTTGTAACGAAAAAACCTTAC 2228
251 AlaValGlnAsp.....GlyGlnGlnGlyValSerSerSerThrSerTh 265
2229 CGACGATTAAGTATGTTGTTTCATTTAGCAAGACGACATCAGAGCGCAT 2278
265 rGlnAspProValValSerPheSerArgAsnThrAlaValGln..... 279
2279 TCAGCTTCCGATCAGCTCATTTAAATCTACAGAGACTTGCACACATC 2328
280 .....Phe 280
2329 AAGCGCAATCTTAGTGCAGCGAGACAGCACTATACGTTACGGCA 2378
281 AspGlyAsnValAlaArgValGlyGlyIleTyrSerTyr..... 294
2379 CGCCACCCAAAGCGCAACCTCAGCTCGTGGCAATGCCAAGCAACAT 2428
295 .....GlyAsnValAlaPheLeuAsnAsnGlyLysThrLeuP 307
2429 TT.....AATCAAGCCACA.....TTA 2445
307 heLeuAsnAsnValAlaSerProValTyrIleAlaAlaGlnProThr 323
2446 AAGCGCAACATCGGCTTCGACAATGCTTCATTTAATCTAAGCAACA 2495
324 AsnGlyGlnAlaSer.....AsnThrSerAspAs 333
2496 CGCCGACAAAGCGCAGTCTGACGTTTCGACAACGCTAAGCAACG 2545
333 nTyrLeuAspGlyGlyAlaIlePheCysLysAsnGlyAlaGlnAlaIa 350
2546 TAAGCATTCGCGACTCAAGCGCAATGTCCTTACCGCAAAATCAGCGG 2595
350 LysSerAsnAsn.....SerGlySerValSerPheAspGlyGlnGlyVal 364
2596 TTCCATTTTGAAGACGCGCTTTACCGGAAATCAGCGGCGGCAAGCA 2645
365 ValPhePheSerSerAsn.....ValAlaAlaGlyLysGln 376
2646 TACGCGATTACCTTAAAGACAGCAATGACGCTGCCGTGGCGACG 2695
376 yGlyAlaIleTyrAlaLysLysLeuSerValAlaAsnGlyProValG 393

```

```

2696 AA.....TTAGCAATTTAAACCTTGACAAAGCCACCATTAACCTCAATCC 2742
393 InPheLeuLysAlaIleAlaAsnAspGlyAlaIleTyrLeuGlnGly 409
2743 GCTTATCCAGACGATCCGCGAGCGCGCAAAACGGCACTGGCGCATGC 2792
410 SerGlyGlnLeuSerLeuSerAlaAspTyrGlyAspIleIlePheAspG 426
2793 GCCGCGCGCGCTTCGCGCGCTTATTCGTTACGGCGCAACTT 2842
426 yAsnLeuLysArgThrAlaLysGln.....AsnA 436
2843 CGCGAAGATCCCGTTTCAACAGCTGACGCTAAGCGGAATTAAGCGT 2892
436 IaAlaAsp.....ValAsnGlyValThrValSerSerGlnAlaIleSer 450
2893 CAGGAAACATTCGCTTATATGTCGAACTTTCGGATACCGCAGCGCA 2942
451 MetGlySer.....GlyGlyLys 456
2943 ATTGAACCTGGCGGAAGTCCGAAGGCATTACACCTTG..... 2982
456 sIleThrThrLeuArgAlaLysAlaGlnGlnIleLeuPheAsnAspR 473
2983 .....GCTGTCAACATACCGCGAAGCAACCCGTA...AGTCTGAGCAA 3024
473 rolleGlnMetAlaAsnGlyAsnAsnGlnProAlaGlnSerSerGlnPro 489
3025 TTGACGATAGTGAAGAAAGACACACACCG.....CTGTCCGAAA 3068
490 LeuLysIleAsnAspGlyGlyGlyTyrThrGlyAspIleValPheAlaAs 506
3069 TCTTAATTTCAACCTG...CAAAAGCAACAGCTGATGCGCGCGCATGGC 3115
506 nGlyAsnSerThrLeuLysGlnAsnValThrIleGlnGlnGlyArg... 521
3116 GTTATCAGCTTATCCGAAAGCGGAGATTCGCTGCATTAATCCGCTC 3165
522 .....IleValLeuArgGlnLysAlaLysLeuSerValAsnSer..... 534
3166 AAGCAACAGAGCTTTCGCAAACTCGGC.....AAGCGGG 3203
535 .....LeuSerGlnThrGlyLysLeuTyrMetGlnAlaGln 547
3204 AGAAACAGAGGCGCGCTTGACGCGCAAAAGCGCAACTT.....GCCG 3247
547 ySerThrLeuAspPheValThrProGlnProProGlnProProAla 564
3248 CCAAACAAAGCGCGGAAAGACACAGCGCA...AGCTTGAAGCGCTG 3294
564 IaAsnGlnLeuIleThrLeuSerAsnLeuHisLeuSerLeuSerLeu 580
3295 ATTGCGCGCGCGCGCAATGCGACGCAAGCAAGCAAGTTCGCGCAAC 3344
581 LeuAlaAsnAsnAlaValThrAsnProThrAsnProThrAlaGlnAs 597
3345 GCGCGCGCAGGCA.....GGCGGGAATAATGCCGCGCATTAAGCGG 3388
597 pSerHisProAlaIleIleGlySerThrThrAlaGlySerValThrIle 614
3389 AG.....GAAGACAAATAACGGGTGCAGCGGATTAAGAC 3423
614 eGlyProIlePhePheGlnAspLeuAspAspThrAlaTyrAspArgTyr 630
3424 ACCGCTTGGGGAAGACGCGGCAAGCGGAACCGCGCGCTACCAACGC 3473
631 AspThrLeuGlySerAsnGlnLysIleAspValLeuLysLeuGlnLeuG 647
3474 CTTCGCCGCGCGCGCGCGCGCGCGCGGATTTGCGCAACCGCAACCC 3523
647 yThrGlnProSerAlaAsnAlaProSerAspLeuThrLeuGlyAsnGln 664
3524 AACCGCAACCCCAACCGCGAGCGGACCTGATCAGCCGTTATGCCAATAC 3573

```



```

505 .....ACAGATGCAGAACTGTTGAGATGACGATT 535
114 userSerAlaThrSerThrSerGluTyrAsnProIleIleGlnArgp 131
536 ATATGATGGG.....TGG 549
131 heLysAspGlyThrPheSerLeuGlyThrLeuValSerGluGlySerLeu 147
550 AAATACGCTGATTTAAATAAATACCTGATCGTTGTAATCGACGACG 599
148 LysIleHisTyrIleAsnGlnSerGlyAsp..... 157
600 CAGCAATATTTGGCGTGTATGATGACGACCAACCAATACCGGAAAGTT 649
158 ..SerLysTyrTrp..... 161
650 CATATCATATTGCAAGCGCATATTCTTGCGTGGTGGCATACCTTT 699
162 .....ThrPhe 163
700 GCACAAATGATCAGTGTGTCGACAGTCACCTAGTAGCGCAAAAT 749
164 ArgArgAsp.....GlyGlyPheThrValAspGlyGly..... 174
750 TAAACATAGCCCATATGTTTACCAACAGAGGCTCATTTGGCGACA 799
175 .....GlyLeuGlyValSerGlySerIleThrThr 186
800 GTGGC.....TCACCAATGTT..... 816
186 erGlyAsnIleAlaIleLeuGlyAsnIleThrSerProGlnIleAsnThr 202
817 .....ATCTATGATGCCCAAAA 833
203 LysAsnIleIleLeuAspThrLysAlaPheGlyGlnTyrAspSerGln.. 218
834 GCAAAATGGTTAATTAAATGGGATTGCGAAAGATGGTTCTATGATGA 883
219 .....SerLeuValGlnTyrValTyrProGlyThrG 229
884 GAAAAAGCAATGGCTTCCAGCTAGTTCGTAAAGATGGTTCTATGATGA 933
229 LysGluGlnAsnGlyIleAsnTyrLeuArgLys..... 239
934 ATCTTGCTGGAGATACCACTTACAGTATCTACGAA.....CCACATGA 977
240 ValArgAlaLysSerGlyGlyThrIleTyrHisGluIleAlaSerAlaG 256
978 AAATGGGAATACTTTTAAAGAC..... 1002
256 nIlnGlyLys.....AsnAspGlnIleSerTyrThrGlyAsnT 270
1003 .....AAATAATGGCGCA..... 1017
270 hIleuThrThrLysLeuMetGlyLeuArgAsnAspGlyAlaMetValLeu 286
1018 .....GGAATAATC.....GATGCCAAACATAAACA 1043
287 ArgArgSerLeuAlaIleGlyThrIleThrAlaAspGlnAsnThrAsnAs 303
1044 CTATTCCTACCTTATAGATTAATAAACAGAACCGTTCAATGTTTAAG 1093
303 nTyrGlySerProThrProMetGlyGlnArgTyrIleAlaLeuGlyAsp. 319
1094 TTTCCTTATCCAGACAGACAGAACCTGTTATCATGCTGACGAGTGGG 1143
320 .....AlaAlaThrGly 323
1144 GTCAACAGTTATCGACCACTGATTAATGGAATAATATTTCCCTTAT 1193
324 .....LeuLysTyrTrl 327
1194 TGCAAGGAAAGTGAATTGATCTTACGACGAATCAACCAAGCGC 1243

```

```

1 ..... 342
327 eLysGlnGlyValTyrAspLeuVal.....GlyAsnTrpAsnSerValA 342
1244 CGGGCGGTTTGTATTTTGAAGGTATTTACGGTCTCGCCATAAAACAC 1293
342 La.....SerIleThrProAspSerPhe 349
1294 GAACGTCGCAAGCGCGGCTTCATATCATGATGATGACGATACCGTTAC 1343
350 ArgSerThrArgLysAlaLeuPheGlyArgSerGlnAspGlnGlyGly 366
1344 TTGGAAAGTAAACGCGCTGGCAACGACCGCTGTCCAAATTCGGCAAG 1393
366 rTrpThrMetProGlyThr.....AsnAla 375
1394 GCACGCTGCTGTCACAGCAAGGGAACCAAGGCTCGTGCAGCGTG 1443
375 laLeuLeuSerValGlnThrGlnAlaAspValAsn.....AsnAla 388
1444 GCGCAGGTAAAGTCATCTTAGATCAGCGCGGACGATCAAGCAAAA 1493
389 GlyAspGlyGln..... 392
1494 ACAACCTTTTACTGAATCGCGTTGTCAGCGC..... 1527
393 .....ThrHisIleGlyTyrAsnSerGlyGlyLysMetSerHisT 406
1528 .....AGGGGCGAGGTGCACACTGAATGCCGATTAATCAG..... 1560
406 YrPheArgGlyLysGlyGlnThrAsnIleAsnThrGlnLysGlyMetGlu 422
423 ValAsnProGlyIleLeuLysLeuValThrAspSerAsnAsnValGlnPh 439
1561 TTCAAACCCCGAC.....AAACTATATTGCGCTTCCGCGGACGCTTT 1604
439 eTyrAlaAsnGlyThrValSerSerIleGlnArgLysLys...PheAspA 455
1605 GGATTTGAACGGGCAATTCGCTTTCACACCGCATTCAAATACCGATG 1654
439 eTyrAlaAsnGlyThrValSerSerIleGlnArgLysLys...PheAspA 455
1655 AAGGGCGCATGATGTCACCAACCAATCAAGAC.....AAAGATTC 1695
455 snGlyLeuValLeuThrGlyAlaArgProAspGlyIleGlnLeuAspAla 471
1696 ACCGTTACCATTTACAGCAATTAAGATATTACTACACCGGCAATTAACA 1745
472 ProThrAlaAlaAspGlyThrLysThrIleLeuThrAlaGlyGlyThrAr 488
1746 CAACTTGATACGAAAAAAGAAATTCGCTACACGCTGTTGGCGACA 1795
488 gaLagIyGlnAsnLysSerTyrValSerIleLysAlaTrpGlyAsn... 503
1796 AAGATGCAACCAAAACGAACGCGGCTCAATCTGAATTTACCAACGGAA 1845
504 .....SerPheAsnAlaSerGlyAsp 510
1846 GAAGCGGATCGCACTTTACTGCTTTCCGCGGCAACAAATTTAAACGGCAA 1895
511 ArgAlaArgGluThrValPhe.....G 518
1896 TATCAGCAAAACGAACGCAACTGTTTTCAGGGGACAGACCGACCGC 1945
518 uValGlyAspGlyGlnGlyPheHisPheTyrSerGlnArgValAlaProA 535
1946 ACGCTACATCATTTAGAGAGCGGTCGTCAAATAATGAGAGT..... 1989
535 laProGlySerThrValGlyProIleGlnLeuAlaGlyValAsnGlyGlyLeu 551
1990 ATCCCAAGAGGAATTCGTGTGGACACGATTTGATCGACCGCATTT 2039
552 LeuThrAlaGlySerIleValAlaSerGlySerIleThrThrGluSer.. 567
2040 TAAAGCGAAACTTCATATTTCAGGGCGGCAACGCGTGGTTCCCGCA 2089

```

```

568 .....SerLeuasnValaasnsllyLeuSerVal.....AsnG 579
2090 ATGTTGCCAAAGTGGAAGCGGATGGCATTTAAGCAATACGCCCAAGCA 2139
      ||||| ||||| ||||| |||||
579 LyglnAlaIysPheIyglYThrAlaAsnAlaLeuArgIleTPrsnAla 595
      ||||| ||||| ||||| |||||
2140 GTTTGCGTGTCCACCGCATCAAAAGCCACAATCTGTACAGCTTCCGA 2189
      ||||| ||||| ||||| |||||
596 GluTyGlyAla.....LlePheArgArgSerGI 605
2190 CTGGAGCGGCTGTGACAAGTTGTACGAAAAAACATTCACGACGATAAG 2239
      : ..... : SerAsnPheTyri 610
605 U..... : ..... :
2240 TGATTCGTCATTTGAGACAAGCCGACATCAGAGCAATGTACGCTTCC 2289
      ||||| : ||||| |||||
610 LeIlePThrTrAsnGlnAsnGlnGlyGluSerGIYAspIle..... 623
2290 GATACGCGTCATTAATATTCACAGACACTTGCACATCCACGCAACTCT 2339
      ||||| : ||||| : |||||
624 ...HisSerSerLeuArgProValArgIleGlyLeuAsnAspGIYAlaVa 639
2340 TAAGTCAGGCGGAGACAGCAGCACTATACGGTTAAGCGCAAGCC.....A 2383
      : ||||| : ||||| : |||||
639 IGIyleuGIyArGAspSerPheIleValAspIlnAsnAsnAlaLeuThrT 656
2384 CCAAAAGCGCAACCTCAGCCCTGCGGGCAAT.....GCC 2418
      ||||| : ||||| : |||||
656 hrIleAsnSerAsnSerArgIleAsnAlaAsnAlaPheArgMetGlnleuGIy 672
2419 CAAGCAACATTTAATTCAGACCCACATTAACGGGCAACATCGGCTTCGA 2468
      ||||| : ||||| : |||||
673 GlnSerThTyrlleAspIlaGluCysThrAspTrValArgProIaGI 689
2469 CAATGGTCATTTAATCTPAGCAACAAGCCGCTACAAAGCGAGTCTGA 2518
      ||||| : ||||| |||||
689 yAlaGIySerPheValSerGlnAsnAsn..... 698
2519 CGCTTTCGACACGCTTAGCA.....AACGTAAAGCATTC 2556
      : ||||| : ||||| : |||||
699 .....GluAsnValArgAlaProPheTyMeAsnIleAsnArgThr 712
2557 ..... : ..... : GCATCAACGGCAA 2570
713 AspThrSerThTyValProIleLeuIySglnArgIyValAlGlnGIyAs 729
2571 T.....GTCTCCCTAGCCGATACGACGAT 2596
      ||||| : ||||| : |||||
729 nSerCysTySerLeuGIyThrLeuIleSerThChIyAspPheArgIle 745
2597 TCATTTTAAAAACAGCCGCTTTACCGGAAAAATCAGGGCGGCAAG... 2643
      ||||| : ||||| : |||||
746 .....HisTyrlsGIuGIyGlnAspAsnGIySerThnGIyProGlnIyAla 761
2644 GATACGCGCATTTACACTTA...AAACACGCGCAATGAGACGTCGCGTGGG 2690
      ||||| : ||||| : |||||
762 AspLeuAlaTrpGlnPheArgArgAspGIySerPheArgSerProAsnIy 778
2691 CACGCAATTA.....GCGCAATTTA..... 2709
      ||||| : ||||| : |||||
778 sIleGIuIleAsnAlaValTrnIleGIyThrAspGIyAsnIleThnGIyG 795
2710 .....AACCTTGACACGCGCAACCTTACATTCATTTCCGCTATGCA 2751
      ||||| : ||||| : |||||
795 LyThnGIyAsnPheAlaAsnLeuAsnThrTrnLeuAsn.....Arg 808
2752 CACGATGCGGCGGCGGCAACCGGCAAGTCGGGACAGATGCGCGCGCG 2801
      : ||||| : ||||| : |||||
809 IySthThrValGIyGIyTrPalagIySerSerValValGIyTrpyrIy 825
2802 CGGTCGCGCGCTCCCATATATCCGTATACGCCGCCCAACTTCGGCAGAA 2851
      ||||| : ||||| : |||||
825 S.....PheAlaTrnValThrIleProGlnSer..... 834

```

[illegible]

```

3604 AGCGTTTCGCGGTACAGGACGAATTTGACCCGCGTTTTCGGAAGACCG 3653
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1091 SerIyIleThrPheMetAlaSpHISglYAspLeuSerAsnSeraSpse 1107
3654 CCGGACGCGCGTTTGACAGGCGGATCCGGGACACCAACACTACCGTT 3703
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1107 rTyTyTyProIle.....ValGlyAlaTySerA 1117
3704 CGCAATATTTCCGCGCTACCGCCACCAACACCGACCTCGGCAATTCGT 3753
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1117 snTyTyTySerAlaGlyTyTyTyGlnThrPheGlnPheGlyTyTyTyGly 1133
3754 ATGCAGAAAAACCTCGGACGCGCGCGCTCGCATCTGTTTTCGCACA 3803
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1134 SerGlyTyThrAlaGly...TrpArgAspGlyIleIleArgIleArgI 1149
3804 CCGGACCGGAAACACTTCGACGAGCGCATCGGCAACCTCGGACGCGCTG 3853
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1149 Y.....AspAsnAlaAsnGlyGlnGlnAlaArgTyTPA 1160
3854 CCCACGCGTCCGCTTTTCGGGCAATACGCG..... 3882
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1160 rg.....PheThrMetAspGlyTyThrLeuAspGlyTyLys 1172
3883 .....ATCGGACAGTTTCGACATCGGCGATCGCGCGCGCGC 3917
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1173 ValLeuLeuProGlnThrGlyAlaPheGlyValAsnThrSerAsnGly 1189
3918 GGGTTTATAGAGCGGCGCTTTCACAGCGCATCGACGCAAAATTCGCGC 3967
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1189 uGlyGlyAsnSerIleThrPheGlyAspSerAspThrGly...IleLysG 1205
3968 GCGCGCTGCTGCATTCAGGCAATTCAGCAAGATACCGCGAGTTTCGCG 4017
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1205 lAsn.....Gly 1207
4018 GGATTTCGATCGAACCACATCGCGCGACGCGCTATTC...GTCCA 4064
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1208 AspGlyLeuLeuAspIleTyAlaAsnSerValGlnValPheArgPheG 1224
4065 AAAGAGCGAT...TACCGATACGAAACGTAATATCGCGACCGCGCGC 4111
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1224 nAsnGlyAspLeuTySerTyIleAsnIleAsn..... 1235
4112 TTGCAATTCACCGCTAC...CGCGCGCGCATTAAGCAGATTAATTCATTC 4158
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1236 ..AlaProAsnValTyIleArgSerAspIleArgLeuLysSerAsnPh 1251
4159 AAACGCGCGCACAC 4173
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1252 LysProIleGlnAsn 1256

```

seq_name: SwissProt_40:AMYH_YEAST

seq_documentation_block:

ID AMYH_YEAST STANDARD: PRT: 1367 AA.

AC P08640; P08068;
 DT 01-AUG-1988 (Rel. 08, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Glucoamylase S1/S2 precursor (EC 3.2.1.3) (Glucoan 1,4-alpha-glucosidase) (1,4-alpha-D-glucan glucohydrolase).
 GN STA1 OR STA2 OR MAU5 OR YIR019C.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 CC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
 NCBI_TaxID=4932;
 RX [1]
 RN
 RP SEQUENCE FROM N.A.
 RC STRAIN=S288C / AB972;
 RA Barrer B.G., Badcock K., Bankier A.T., Bowman S., Brown D., Churcher C.M., Connor R., Copsey T., Dear S., Devlin K., Fraser A., Gentles S., Hamlyn N., Horsnell T.S., Hunt S., Jagers K., Jones M.,

```

RA Louis E., Lye G., Moule S., Moule T., Odell C., Pearson D.,
RA Rajandream M.A., Riles L., Rowley N., Skelton J., Smith V.,
RA Walsh S.V., Whitehead S.;
RL Submitted (DEC-1994) to the EMBL/Genbank/DBJ databases.
RN
RP SEQUENCE OF 1-242 AND 762-1331 FROM N.A.
RX MEDLINE=87194600; PubMed=3106330;
RX Yamashita I., Nakamura M., Fukui S.;
RT "Gene fusion is a possible mechanism underlying the evolution of
RT SPAL";
RL J. Bacteriol. 169:2142-2149(1987).
RN
RP SEQUENCE OF 1-31 FROM N.A.
RC STRAIN=SPX101-1C;
RX MEDLINE=89031230; PubMed=3141213;
RA Pardo J.M., Ianez E., Zalacain M., Claros M.G., Jimenez A.;
RT "Similar short elements in the 5' regions of the STA2 and SGA genes
RT from Saccharomyces cerevisiae.";
RL FEBS Lett. 239:179-184(1988).
CC -1- CATALYTIC ACTIVITY: Hydrolysis of terminal 1,4-linked alpha-D-
CC glucose residues successively from non-reducing ends of the chains
CC with release of beta-D-glucose.
CC -1- SIMILARITY: TO S.POMBE SPOC215.13.
CC -1- SIMILARITY: SOME, TO S.POMBE SPOC285.13C.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sdb.ch/announce/
CC or send an email to license@sdb.ch).
CC
DR EMBL: Z38061; CAA86176.1; -
DR EMBL: M16164; AAA35014.1; -
DR EMBL: M16165; AAA35015.1; -
DR EMBL: X13857; CAA32069.1; -
DR PIR: B26877; B26877.
DR PIR: A26877; A26877.
DR PIR: S48478; S48478.
DR SGD: S0001458; MOC1.
KW Hydrolyase; Glycosidase; Polysaccharide degradation; Glycoprotein;
KW Signal; Multigene family.
FT SIGNAL 1 21 POTENTIAL.
FT CHAIN 22 1367 GLUCOAMYLASE S1/S2.
FT DOMAIN 210 1367 SER/THR-RICH.
FT CARBOHYD 817 847 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 874 874 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 1367 AA: 136110 MW: 91C00E2DBD61AA9D CRC64;

```

alignment_scores:

Quality: 199.00 Length: 999
 Ratio: 0.423 Gaps: 48
 Percent Similarity: 47.047 Percent Identity: 20.220

alignment_block:

US-09-303-518D-653 x AMYH_YEAST ..

Align seg 1/1 to: AMYH_YEAST from: 1 to: 1367

```

1510 ATCGGCTTGGTCACGCGGAGGCGGCGTGCACATAATCCGATTAATCA 1559
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
99 IleGlyValThrGlyProLysGlyTyThrValGlnLeuTyrcGlyTyAsnG 115
1560 GTTCAAC.....CCGACAACTGATATTGGCGCTTC... 1591
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
115 u.AsnThrTyTrLeuIleAspAsnProThrAspPheThrAlaThrPheGln 131
1592 .....GCGCGGACGCTTGGATTTGAACGGG 1617
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
132 ValTyAlaThrGlnAspValAlaAsnSerCysGlnValTrp...MetProAsn 147

```



```
1618 CATTCGCTTTCGTTCCACCGCATTTCAAAATACCGATGAAGGCGCATGAT 1667
      :::::::::::::: ::::::::::::::
148 PheGlnIleGlnPheGlnIuTrLeuGlnIySerAlaIaIa..... 160
1668 TGCAACACCATCAAGACAAGAAATCCACCGTTACCATTAACAGGCAATA 1717
      :::::::::::::: ::::::::::::::
161 .....GlnTrAlaSerSerTrpGlnTrpGlyThrThrSerP 173
1718 AAGATATTACTACACCGCATTAACACAACTGGATAGCAAAAAAGAA 1767
      :::::::::::::: ::::::::::::::
173 heaSpLeuSerThrGlyCysAsnAsnTrpAspAsnGlnIyHisSerGln 189
1768 ATTGCCTACACAGGCTTGTTGGCGAAGAAATGCAACCAAAACGAACGG 1817
      :::::::::::::: ::::::::::::::
190 ThrAspPheProGlyPheTrpPheIleAspCys..AspAsnAsnCysG 206
1818 GCGGCTCATCTGAATTACCAACCGGAAGACGGATCGACTTACTGCG 1867
      :::::::::::::: ::::::::::::::
206 LysIlyThrLysSerSerThr..... 212
1868 TTTCGCGCGAACAATTTAAACGGCAATATCAGCGAACAACGGAACAA 1917
      :::::::::::::: ::::::::::::::
213 .....ThrThrSerSerThrSerGluSerThr 222
1918 CTGTTTTCAGCGGCGACACCGACCGACGCGCTTCAATCATTTAGGAAG 1967
      :::::::::::::: ::::::::::::::
222 rThrThrSerSerThrSerGluSerSerThrThrThr..... 234
1968 CCGGTGGTCAAAAATGGAAGTATCCCAACAGAGAATCGTGTGGACA 2017
      :::::::::::::: ::::::::::::::
235 .....SerSerThrSerGluSerSer.....Thr 242
2018 ACGATTGATCGACGCGACATTTAAAGCGGAACCTCCATATTCAGGCG 2067
      :::::::::::::: ::::::::::::::
243 ThrThrSerSerThrSerLysSerThrSerSerSerThr..ThrAlaP 259
2068 GGACAACGCGTGTTCGCCGATGTGCCAAGTGAAGCGCATTTGGCA 2117
      :::::::::::::: ::::::::::::::
259 roAlaIaThrProThrThrThrSerCysThrLysGlu..... 270
2118 TTTAAGCATACAGCGCCAGACAGTTTGGTGGCA..... 2154
      :::::::::::::: ::::::::::::::
271 ...LysProThrProThrThrThrSerCys..ThrLysGluLysProT 286
2155 .....CCGATCAAAAGCCACACATCTGTACAGCTTCGACTGACGGGT 2199
      :::::::::::::: ::::::::::::::
286 hrProProHisHisAspThrThrProCysThrLysLysThrThrThr 302
2200 CTGACAACTGTACCGAAGAAACATTAACCGACGATTAAGTATGCTTC 2249
      :::::::::::::: ::::::::::::::
303 SerLysThrCysThrLysThrLysThrThrPro.....ValProThr 316
2250 ATTGAGCAAGACGACATCAGAGCAATGTACGCTTGGCATTCAGCGTC 2299
      :::::::::::::: ::::::::::::::
316 rProSerSerSer..ThrThrGluSerSerSerAla...ProValProThr 331
2300 ATTTAAATCTCACAGACTTGCACACTCAACGAGCAATCTTAGTCAGAGC 2349
      :::::::::::::: ::::::::::::::
332 Pro...SerSerSerThrThrGluSerSerSerAlaProValThrSerSe 347
2350 GGAGACAGCACTATACGTTACGGGCAACGCCACCCCAAAACGGACACT 2399
      :::::::::::::: ::::::::::::::
347 ThrThrThrGluSerSerSerAlaProValProThrProSerSerSerThr 364
2400 CAGCCTCGTGGCAATGCCCAAGCAACATTTAATCAACGCCACATTAACG 2449
      :::::::::::::: ::::::::::::::
364 hrGluSerSerSerAlaPro..... 370
2450 GCAACACATCGGCTTGGACATGCTTCAATTAATCAAGACACAGCC 2499
      :::::::::::::: ::::::::::::::
371 .....ValThrSerSerThrThrGlu 377
2500 GTACAAAACGGCAGTCTGACGCTTCCGACAAACGCTAAGCAACGTAAG 2549
```

```
377 uSerSerSerAla.....ProValThrSerSerThrThr.... 388
      :::::::::::::: ::::::::::::::
2550 CCATTCGCGCATCAACGGCAATGTCTCCCTAGCCGATTAAGCAGTATTC 2599
      :::::::::::::: ::::::::::::::
389 .....GluSerSerSerAlaProValProThrProSerSerSer 401
2600 ATTTGAAACACGCGCTTTACCG.....GAAATATACG 2634
      :::::::::::::: ::::::::::::::
402 ThrThrGluSerSerSerAlaProValThrSerSerThrThrThrGluSer 418
2635 GCGCGCAAGATPACGGCATTAACCTTAAAGACGCGAATGACCGCTGCC 2684
      :::::::::::::: ::::::::::::::
418 rSerAlaProValThrSerSerThrThrThrGluSerSerSerAla..... 432
2685 GTGCGGACGGAATTAGCAATTTAAACCTTGACACGCGCAATTAAC 2734
      :::::::::::::: ::::::::::::::
433 .....ProValThr 435
2735 TCAATTCGCGCTATCAGACAGATGCGGACGCGCAACCGGACGTGCG 2784
      :::::::::::::: ::::::::::::::
436 Ser.....SerThrThrGluSerSerSerAlaProValThrSe 448
2785 GCAGATGCGCGCGCGCGCGCTTCGCGCGCTTCCCTATTATCCGTTACGCC 2834
      :::::::::::::: ::::::::::::::
448 rSerThrThrGluSerSerSerAlaProValPro..... 459
2835 GCCAATCTGCGCAGATCCGCTTCAACACGCTGACGTAACGCGCAAT 2884
      :::::::::::::: ::::::::::::::
460 .....ThrProSer 462
2885 TGAACGCTCAGGAACAATTCGCTTATGTGCGAAGCTTTCGGGTACCGC 2934
      :::::::::::::: ::::::::::::::
463 SerSerThrThrThrGluSerSerSerAlaProValThrSerSerThrThrGlu 479
2935 AGCGCAAAATTTGAAGCTGGCGGAAGTTCGGAAGCACTTACACCTTGGC 2984
      :::::::::::::: ::::::::::::::
479 uSerSerSer.....AlaProValPro.....ThrProSer 490
2985 TGTCACAAATACCGCGCAACGCCGTAAGTCTGACAAATTTGACGCTAG 3034
      :::::::::::::: ::::::::::::::
490 erSerThrThrGluSerSerSerSerAlaProValThrSer..... 503
3035 TGAAGAGAAAACACACACACCGCTGTCCGAAATCTTAATTCACCCCTG 3084
      :::::::::::::: ::::::::::::::
504 .....ThrThrGluSerSerSerAlaProValProThrProSe 516
3085 CAAACGACACAGTGCATGCCGCGGCATGGCGTTATCAGCTTATCCGCA 3134
      :::::::::::::: ::::::::::::::
516 rSerSerThrThrGluSerSer.....SerAlaProAlaP 528
3135 AGACGGGAGATTCGCGCTGATATTCGGTCAAAAGAACAGAGACTTTCG 3184
      :::::::::::::: ::::::::::::::
528 roThrProSerSerSer.....ThrThrGluSerSerSerAlaPro 541
3185 ACAAACTCGCAAGGCGGGAAGAAACAGAGGCGCCCTGACGCGAAGAACAG 3234
      :::::::::::::: ::::::::::::::
542 ValThrSerSerThrThrThrGluSerSerSerAlaPro..... 553
3235 GCACACTTGGCG...CCAACACACAGCGGGAAGAAACACACAGCGCAAG 3281
      :::::::::::::: ::::::::::::::
554 .....ValProThrProSerSerSerThrThrGluSerSerSerThrP 568
3282 CCTTGAGCGCGCTGATTTGCGCGCGCGGCGCAATGCCACCGAAGGAGAAA 3331
      :::::::::::::: ::::::::::::::
568 roValThrSerSerThrThrGluSerSerSerAlaPro..... 580
3332 GTGTTGCCGAACGCGCGCGCAGGACGCGGGAAGAAATGCCGCGCATTAAG 3381
      :::::::::::::: ::::::::::::::
580 ..... 580
3382 CAGCGGAGGAGAGAGAAAAACGGGTGACGCGGATTAAGACACGCGCTT 3431
      :::::::::::::: ::::::::::::::
```

```

581 .....ValProthrPro.... 584
3432 GCGGAACAGCGCGAAGCGAAACCGCGCGCTACACCGCCTTCCCC 3481
      ::::: ::::: ||| ||:::|
585 .....SerSerSerThrThrGluserSerSerAlaProValProthrPro 599
      ::::: ::::: :::::
3482 GCGCGCGCGCGCGCGCGGATTTGCGCAACCGCGACCGCCACCGCA 3531
      ::::: ::::: :::::
600 SerSerSerThrThrGluser.....SerSerAlaProAlaProth 613
3532 CCGCAACCGCGCGCGCTGATCAGCGCTTATGCCAATAGCGGTTGAG 3581
      ::::: ||:::| ||:::|
613 rProSerSerSerThrThrGluserSerSerAlaPro..... 626
3582 TGAATTTTCGCGCAGCTCAACAGCGTTTCCCGTACAGAGCAATGG 3631
      ::::: ::::: :::::
626 alThrSerSerThrThrGluserSerSerAlaProValProthr..... 640
3632 ACCGCGTGTTCGCGAGACCGCGCGACCGCGTTGGCAACAGCGCATC 3681
      ::::: ::::: :::::
641 .....ProSerSerSerThrThrGluserSerSerAlaProVa 653
3682 CCGGACACCA...AACACTACGTTGCGAGATTCCGCGCTACCGCA 3728
      ||:::| :::::| :::::|
653 lProthrProSerSerSerThrThrGluserSerSerAlaProVal.... 668
3729 ACAAACGAGCTGCGCAATCGTATGCGAGAAACCTCGCGACCGGCG 3778
      ||:::| :::::| :::::|
669 .....ProthrProSerSerSerThrThrGluserSerSerAlaProVal 683
3779 GCGTCGCGACCTGTTTTCGCAACACCGCGAGAAACCTTGCACGAC 3828
      :::::| :::::| :::::|
684 ThrSerSerThrThrGluserSerSerAlaProValThrSerSerThrTh 700
3829 GGCATGCGACACGCGCGACGCGTGGCGCGGCTTGGCGCAATA 3878
      | ||:::| :::::| ||| ||:::|
700 rGluserSerSerAla...ProValProthr...ProSerSerSerThrTh 715
3879 CGGACATG.....GCAAGTTGCAATCGGCA 3904
      || ||| :::::| :::::|
715 hrGluserSerSerAlaProValProthrProSerSerSerThrThrGlTh 731
3905 TCAGCGCGCGCGGCTTTAGTAGCGGACCTTCAGACGGCATCAGA 3954
      ||:::| ||| ||| :::::|
732 SerSerSerAlaProValProthrProSerSerSerThrThrGluser.. 747
3955 GCGAAATCGCGCGCGCTGCTGCATTACGCGATTACAGCAAGATACCG 4004
      ||:::| ||:::| :::::|
748 .....SerSerAlaProValThrSerSerThrThrGluserSerSerAla 762
4005 CCGAGGTTTCGCGCGATTCGCGATCGAACCGCATCGCGGACGCGCT 4054
      || ||| :::::| :::::|
762 lAProValProthrProSerSerSerThrThrGluserSerSerAlaPro 778
4055 ATTTCGTCCAAAAAGCGATTCGCGATTCGCAAAAGCTCAATTCGCGCC 4104
      ::::: ::::: ||| ||:::|
779 ValProthrProSerSerSerThr...ThrGluserSerSerAlaProVa 794
4105 CCGGCGCTTCATTCACCGCTACCGCGCGGCGCATTAAGCAGATT.... 4150
      ::::: ||:::| :::::|
794 lProthrProSerSerSerThrThrGluserSerValAlaProValProT 811
4151 .....ATTTCATCAACCGCGCGACACATTCATCAGCGCTT... 4189
      ||:::| ||:::| :::::|
811 hrProSerSerSerSerAlaThrThrSerSerAlaProSerSerThrPro 827
4190 .....ATTGAGCGCTGCTTATACCGCGCGCTTCC 4221
828 rProSerSerSerThrThrGluserSerSerValProValProthrProSe 844
4222 GCGAAATCGCGAGCGCGCTCAATACCGCGCTATTGGCGCAGAGATTTCG 4271
      ::::: ||:::| :::::|
844 rSerThrThrThrGluserSerSerAlaPro.....ValSers 856

```

```

4272 CAAAACCGCGAGTGGGATGGGCG.....TAACCGCGAAATCAAG 4315
      ::::: ::::: ||| ||:::|
856 rSerThrThrThrGluserSerValAlaProValProthrProSerSerSer 872
4316 .....GTTTCAGCGCTGCTCCCTCCAGCGCTGCG 4342
      ::::: ||:::| :::::|
873 SerAlaThrThrSerSerAlaProSerSerSerAlaPro 884
seq_name: SwissProt_40: HXA3_HAEIN

```

```

seq_documentation_block:
ID HXA3_HAEIN STANDARD; PRT; 917 AA.
AC P43535;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Heme/hemopexin-binding protein precursor (Heme:hemopexin utilization
DE protein A).
GN HXA.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OX Haemophilus.
NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NTHI N182;
RX MEDLINE=9511556; PubMed=7815944;
RA Cope L.D., Thomas S.E., Latimer J.L., Slaughter C.A.,
RA Mueller-Eberhard U., Hansen E.J.,
RT "The 100 kDa hem:hemopexin-binding protein of Haemophilus
RT influenzae: structure and localization."
RL Mol. Microbiol. 13:863-873(1994).
CC -1- FUNCTION: BINDS HEME/HEMOPEXIN COMPLEXES.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
CC DR EMBL; 008349; AAA/4139.1; -.
KW Transport; Signal; Repeat.
FT CHAIN 1..21
FT SIGNAL 1..21
FT DOMAIN 22..38
FT REPEAT 24..28
FT REPEAT 29..33
FT REPEAT 34..38
FT DOMAIN 111..668
FT REPEAT 111..116
FT REPEAT 203..208
FT REPEAT 277..282
FT REPEAT 399..404
FT REPEAT 629..634
FT REPEAT 663..668
FT DOMAIN 159..170
FT REPEAT 159..164
FT REPEAT 165..170
SQ SEQUENCE 917 AA; 100148 MW; 7FF3BBB8C046539D CRC64;

```

alignment_scores:

```

Quality: 198.00 Length: 1129
Ratio: 0.415 Gaps: 55
Percent Similarity: 42.250 Percent Identity: 19.309

```

alignment_block:

```

US-09-303-518D-653 x HXA3_HAEIN ..

```

Align seg 1/1 to: HXA3_HAEIN from: 1 to: 917

```

442 GAGGAGACTACGAGCCATTCATGCGGC.....GATTATCATATGCC 485
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
21 AlasethrGlnGlyLeuProGlnGlyLeuProGlnGlnGlyLeuP 37
486 GCGTTGGCAAAATTTGTCACAGATGCAAGACCTGTGAGATGACCACT 535
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
37 oGlnGlnAspLysValValValGlyGlnAlaIthrPheAspLysThr 52
536 AATATGATGCGTGGAATACGCTGAT.....TTAAATTAATACCT 576
53 .....|leAlaAspLysMetThrIleAsnGlnIthrSer 63
577 GATCGTTGCAATGCGACGACGACAGACAATATTCGCGGTGATGANGA 626
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
64 AspLysValGlnIleAsn.....TTPHISer...PheAs 74
627 CGAACCCATAACCGCGAAAGTTCAATATTCATTCAGC..... 666
|:|||||:|||||:|||||:|||||:|||||:|||||:
74 pIleGlyGlnAsnLysGluValGluPheLysGlnProSerGluAsnSerV 91
667 ..GCATATCTTGCTCGTGGTGGCATACCTTTGCCAAATGCA... 711
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
91 alAlaIerYAsnArGValThrGlyGlyAsnAlaSerGlnIleGlnGlyLys 107
712 ...TCAGTGGTGCGACAGTCACTAGTAGTACGAAAAAATTAACATAG 758
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
108 LeuThrAlaAsnGlyLysValIlyrLeuAlaAsnPro..... 119
759 CCCATATGCTTTTTCACACAGAGGCTCATTTGGCGACAGTGGCTCAC 808
119 ..... 119
809 CAATGTTATCTATGATGCCAAAAGCAAAAGTGTATTAATGCGGTA 858
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
120 .....AsnGlyVal 122
859 TTGCAACAGCGACACCCCTATATAGAAAAGCATGCTTCACACTAGT 908
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
123 IleIleThrLysGlyAlaGluIleAsnValAla...GlyLeuLeuAlaThr 138
909 TCGTAAAGATTTGTTCTATGATGAATTTTGTGCGAGATACCCATTGAG 958
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
138 rThrLysAsp.....LeuGluGlnIle..... 145
959 TATTTACAGAACACATCAAAATGSGAAATACTTTTAAAGACATATAT 1008
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
146 .....SerGluAsnSer 149
1009 AATGGCGAGAGAAATGATGCGCAACATTAACACTATTCCTACCTTA 1058
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
150 AsnSer.....TyrGlnPheThrAr 156
1059 TAGATTAATAACACGACCGTT.....CAATGTTTAAATGTTT 1096
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
156 gLysThrLysAspArgGlnValValLysGluGlyGlnValIleAsn.... 171
1097 CTTTATCGAGACACAGACAGAACTGTTTATCATGCTGCAGGTGGGTC 1146
172 .....GluGlyGlnIle 175
1147 AACAGTTATCGACAGACTGATATATGAGAAAATATTTCTTATATGA 1196
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
176 LysAlaLysAspPheValValLeuAsnGlyLysGluVal.....IleAs 190
1197 CAAGAGAAAGGTGATTTGATCTTACAGACATCAACCAAGCGCGG 1246
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
190 nLysGlyAsnIleAsnValGluLysAsnSerThrIleAsn.....G 204
1247 GCGGTTTGATTTTGAAGGT.....AATTTACGCTCTCCCTTAAAAAC 1290
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
204 lLysValIlyrLeuSerSerGlyTyrAsnPheThrPheThrLeuSerAsp 220

```

```

1291 AACCAAACTGTGCGAAGCGCGCGCTTCATATCAGTATGCGACTACCT 1340
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
221 Ser.....GlyIleSerValIleLeuIleAsnThr..... 231
1341 TACTTGGAAAGTAAACGCG...GTGCGCAACGACCGCGCTGCCAAATCG 1387
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
232 .....ValArgGlyIleValGlnAsnGlnGlySerIleLysAlaG 245
1388 GCAAGGACGCGCTGCTGTTCAAGCCAAAGGGGAAAAACCAAGGCTCGTC 1437
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
245 lylGluIleThrLeuSerAlaLysGlyArgLysGluAlaLeuAspSerLeu 261
1438 AGCGTGGCGCGGTAAAGTCATCTTATGATCAGAGCGGACGATCAAG 1487
262 ValMetAsnAsnGlyValLeuGluAlaThrLysValSerAsnArgAsnG 278
1488 CAAAAAACAGCCCTTATGTAATCGGCTTGTCACGCGCGGAGGACG 1537
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
278 Lys.....ValValLeuSerAlaAspAsnV 287
1538 TGCACGTGATCCGATATACGTTCAACCCGACAAACTCTATTTGCGC 1587
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
287 alGlnLeuAsnAsnGlnLysSerAsnIleLysGlyGluIleValAsnPheGly 303
1587 ..... 1587
304 ThrGluValIthrSerAsnGluAspLysLysLeuLysIleThrSerLysTh 320
1588 .....TTTCGCGCGGACGTTTGG 1606
320 rGlySerLysValIthrSerProLysIleAsnPheLysGlyLysSerVal 337
1607 ATTGAAACGG.....CATTCGCTTTCG 1629
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
337 snIleAsnGlyAsnPheGlyArgGluAsnSerGlyThrHisTyrAsnGlu 353
1630 TTCACCGCATTCATAATACCGAT.....GAAGGGCGAT 1664
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
354 GluArgLysThrLeuAsnThrGluValAsnIleAspValProGlyAlaG 370
1665 GATTGTCAACCCACATCAAGCAAAAGATCCACCGTTACC..... 1704
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
370 uAsnIleArgIleAlaAspAspLysAspAsnThrGluThrAspSerPhe 387
1705 .....ATTACAGCAATTAAGTATTACT 1728
387 leGlnThrGluAlaLeuSerSerLeuLeuAlaAsnGlyLysValAsn 403
1729 ACAACGCAATTAACAACAACACTGGATAGCAAAAAAGAAATTCCTACAA 1778
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
404 LeuLysGlyAsnAspValAsnIleSerGly.....AsnIleAsnIleAs 418
1779 CGGTGGTTGGCGAGAAAGATGCAACCAAA...ACGAAC...GGCGCGC 1822
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
418 pSerPheArgGlyThrAspSerLeuLeuLysLeuThrAsnGlnGlyHis 435
1823 TCAATCTGAATTACCAACCGGAAGACGATGCGACTTTACTG..... 1866
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
435 leAsnIleAspHisAlaAspIleAsnSerLysGlyArgLeuPheValAl 451
1867 .....CTTTCGCGGAGCAACAATTTAAACGCAATATTCGCAACAAA 1910
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
452 ThrSerLeuGlnAsnAspValAspPheLysSerAsnIleThrIleThrAs 468
1911 CGGCAAACTGTTTTCAGCGGACGACCGACACCGCTACATCATAT 1960
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
468 pSerLysIle.....Asp 473
1961 TAGGAAGCGGCTGCTAAAAATGGAAGGTATCCCAAGAGAAATCGTG 2010
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
473 euGlyAsnGly.....AlaMetGlyLeuGlyArgSerVal 484

```

```
2011 TGGACACGATGATGATCGACCGACATTTAAAGCGAA..... 2049
      :::::|||||
485 AsnGluAspSerPyrAlaThrArgTrpLysArgAlaGluLysSerGlnAr 501
2049 ..... 2049
501 gLysLysPheAsnValAspMetArgAsnValValPheAsnGlnValAspG 518
2049 ..... 2049
518 luvAlIleuValIleGlyIleGlyPheGluLysValAsnLeuAspLysIleVal 534
2050 ..... AACTTCATATTACGGGCGGACAAAGCGGTGCTTC 2084
      |||||:::|||||
535 AlaThrGlyGlnThrAsnPheYrIleAspGlyGly.....ValSe 548
2085 CGCGAATGTTGCCAAAGTGAAAGCGATTGGCATTTAAGCAATCAGCCG 2134
      |||||:::|||||
548 rArgAsnAsnSerArgTrpGlu..... 555
2135 AACGATTTTCGGTTCGACCGCATCAAGCGACATCTGTACA... 2181
      :::::|||||
556 .....TyrGlyValIleuAspLeuAspLysPyrArgThrGlnLeuSerGlu 569
2182 .....CGTTCGGACTGG..... 2193
570 LeuGluGlnGlyArgArgArgTrpArgTrpTyrArgAspLeuAspLeuAs 586
2194 ..... 2194
586 pMetAsnLysAlaTyrLeuTyrArgPheAspLeuPheAlaLysAsnAsnS 603
2195 CGGGTCTGCAAGTGT.....ACCGAAAAACCATTAACCGAC 2232
      ::|||:::|||||
603 erGlyArgSerThrIleLysGlyThrGluIleAsnIleSerAsnSerAsn 619
2233 .....GATTAAGTATTGC 2246
620 IleAsnLeuLysAsnGlyPheValIleLeuLeuAlaGluAsnIleLysLe 636
2247 TTTATGAGACAGACCGACATC.....AGAGCAATGTCA 2281
      ::|||:::|||||
636 uAspAsnSerLysValAspLleThrPheAspLysAspAsnSerGlnAspT 653
2282 GCTTCGCGATCAGCGCTCATTTAATTCACAGGA...CTTGCCACAC 2328
      ::|||:::|||||
653 hIreuAlaGlnThrAsnArgLeuGlyMetAsnGlyLysValSerMetIle 669
2329 AACGCAATCTTAGTCAGGCGGAGACACGACATATACGTTACGGCGCA 2378
      |||||:::|||||
670 AsnSerHisIleLysIleValGlyAspArgLysGluGlyIleSerProth 686
2379 CGCGCAACCAAAAGCGACACTC.....AGCCTCGGGCGAATGCC 2419
686 rGlyThrTyrAlaThrMetSerLeuIleGlyIleLeuIleGlyLys 703
2420 AACCAACATTTAATCAAGC.....ACATTTAAAGCGC 2451
      ::|||:::|||||
703 erSerIlePheValLysSerHisGlnGlyTyrThrPheLysThrAspGly 719
2452 AACACATCG...GCTTCGACAAATGCTTCATTTAATCTAAGCAACAGC 2498
      |||||:::|||||
720 AspThrLysIleAlaGlyLysAsnSerLysGluAspLeuLysIleThrAl 736
2499 CGTACAA.....AACGGAGTGC 2515
736 alIleAsnThrGlyIleArgAlaAlaGluGluValLeuIleAsnGlyAla 753
2516 TGACGCTTCGACGACGTAAGCAAGCAATTCGCGACACTCAAC 2565
      ||:::|||||
753 euGlySerIleAspAsn...AspAlaAsnIleAlaAsnMetAlaPheThr 768
2566 GGCAATGTCTCCTAGCCGATAAAGCATATTCATTTGAAAAACAGCGC 2615
```

```
769 lIleGlyAspSerAlaAsnThrLysThr.....IleGluAsnAlaAs 783
      ::|||:::|||||
2616 CTTTACCGCGAAAAATCAAGCGCGCAAGGATACGGCATTTACACTTAAAG 2665
      ::|||:::|||||
783 pIleThrAlaLeuAlaProAsnGlyGlyThrAlaTyrLysSerSerLysA 800
2666 ACAGGAAATGAGACGCTGCGCTCGGCAAGCAATTAAGCAATTTAACTT 2715
      |||||
800 spValGlu..... 802
2716 GACAAGCGCACCATTAACATCAATTCGCTATCGACAGATCGCGAGG 2765
      ::|||:::|||||
803 .....IleGluValLysProAsnSerAspHe..... 811
2766 CGCGCAAAACCGGACGTGCGGAGATGGCGCGCGCGCGCTTCGCGCGGT 2815
      ::|||:::|||||
812 .....ThrPheGluLeuProArg..... 818
2816 CCTATTATTCGTTACGCGCGCACTTCGGCAATCCGTTCAACACG 2865
      |||||:::|||||
819 .....GluLysAsnLeuAsnGln 824
2866 CTGACGGTAAAGCGC.....AAATGAACGCTCAGGAGACATTCG 2906
      ::|||:::|||||
825 ThrLysIleAsnGlyAlaSerThrLysLeuSerGluArgGlyPheAlaAr 841
2907 CTTTATGCGGAACTCTTCGCTACCGCAGCGGCAATTAAGCTGCGCG 2956
      |||||:::|||||
841 gLeuTyrAspLysIleAsnGlyValArgAlaSerAsnLeu..... 854
2957 AAAGTTCGAAGCAGCATTAACCTTGCTGCTCAACAATACCGGCAAGAA 3006
854 ..... 854
3007 CCCGTAAGTCTGACGAATTAAGCGTATGAGAGAAAGAC..... 3048
      |||||:::|||||
855 .....SerAlaGluGlnLeuAsnValThrAspAlaSerGluLysIleI 869
3049 AACACACGCGTCCGGAATCTTAATTC.....ACCTGCAAAACG 3091
      |||||:::|||||
869 eAsnThrLysLeuValSerSerLeuAspValGluLysLeuValSerValA 886
3092 AACACGTCGATCGCGCGCATGGCTTATCCGCAAGAGCGC 3141
      |||||:::|||||
886 lAlaLysAspAlaGly..... 891
3142 GAGTTCCGCTGCAATTCGCTCAAGACAAAGACTTCCGCAAACT 3191
      |||||:::|||||
892 .....AsnGlyCysGluGlnGlnPheGlyAspLysG 903
3192 CGGC.....AAGCGGAGAAACAGAGCGC 3216
      ::|||:::|||||
903 yAsnAsnThrLysValSerValGlyGluLeuGluAla 915
seq_name: SwissProt_40:MAP4_MOUSE
seq_documentation_block:
ID MAP4_MOUSE STANDARD: PRT: 1125 AA.
AC P27546;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Microtubule-associated protein 4 (MAP 4).
GN MAP4 OR MAP4P.
OS Mus musculus (mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92042100; PubMed=1718985;
RA West R.R., Tendarge K.M., Olmsted J.B.;
```

"A model for microtubule-associated protein 4 structure. Domains defined by comparisons of human, mouse, and bovine sequences.";
 J. Biol. Chem. 266:21886-21896(1991).
 CC -1- FUNCTION: NON-NEURONAL MICROTUBULE-ASSOCIATED PROTEIN. PROMOTES MICROTUBULE ASSEMBLY.
 CC -1- TISSUE SPECIFICITY: TESTIS, STRIATED AND CARDIAC MUSCLE.
 CC -1- PTM: PHOSPHORYLATION OF THE PRO-RICH REGION IN THE C-TERMINUS NEGATIVELY REGULATES MAP-4 ACTIVITY TO PROMOTE MICROTUBULE ASSEMBLY.
 CC -1- SIMILARITY: CONTAINS 3 TAU/MAP REPEATS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; M72414; AAL6372.1; -.
 CC PIR; B41206; B41206.
 CC MGD; MGI:97178; Map4.
 CC InterPro: IPR001084; Tubulin-bind.
 CC Pfam; PF00418; tubulin-binding; 4.
 CC PROSITE; PS00229; TAU_MAP; 3.
 CC Microcubules; Repeat; Phosphorylation.
 CC REPEAT 896 926 TAU/MAP MOTIF.
 CC REPEAT 965 995 TAU/MAP MOTIF.
 CC REPEAT 996 1026 TAU/MAP MOTIF.
 CC SEQUENCE 1125 AA; 117675 MW; 73047432A329A1D CRC64;

alignment_scores:

Quality: 198.00 Length: 1022
 Ratio: 0.406 Gaps: 55
 Percent Similarity: 47.750 Percent Identity: 22.505

alignment_block:

us-09-303-518d-653 x MAP4_MOUSE ..

Align seg 1/1 to: MAP4_MOUSE from: 1 to: 1125

```

1666 ATTGTCACACCATCATGACAAAGAACATCCAGTACCATTCAGGCGAA 1715
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
37 IIEVAlGlyIuThrValGluIuThrGluPheIleProIeuLeuAspG1 53
: : : : : : : : : : : : : : : : : : : : : : : : :
1716 TAAAGATTTACTACACGCGCAATAACACACACTTGATAGCAAAAAG 1765
: : : : : : : : : : : : : : : : : : : : : : : : :
53 YAspGlu.....LysThrGlyAsnSer.....GluSerLysLysL 65
1766 AATTCGCTACAAACGGTTGGTTGGCGAAGAAATGCAACCAAAAGAAC 1815
: : : : : : : : : : : : : : : : : : : : : : : : :
65 YsProCysLeu.....AspThrSerGlnValGlu 74
1816 GGGCGGCTCATCTGATTCACCAACGGAAGAGCGATCGCACTTTACT 1865
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
75 Gly.....IleProSerSerLysProThrLeuLe 84
1866 GGTTCGCGCGAACAATTTAAAGCGCAATATCAGCGCAAAACAAACGCA 1915
| : : : : : : : : : : : : : : : : : : : : : : : :
84 uAlaAsnGlyAspHisGlyMetGluGlyAsnAsnThr..... 96
1916 AACTGTTTTCAGCGCGACAGCGACGCGCGCTACCATCATTTA... 1962
: : : : : : : : : : : : : : : : : : : : : : : : :
97 .....AlaGlySerProThrAspPheLeuGluGluArgValAsp 109
1963 .....GGAAGCGGGTGGTCAAAAATGAGAGATATCCC 1994
: : : : : : : : : : : : : : : : : : : : : : : : :
110 TyrProAspTyrGlnSerSerGlnAsnTrpProGluAspAlaSerPheC 126
1995 A...CAAGAGAAATCGTGTGGACAAACGATGATGACGCGCAACATTTA 2041
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
126 sPheGlnProGlnGlnValLeuAspThrAsp.....G 137

```

```

2042 AAGGGGAAACTTC.....CATATTCAGGCGGACACAGC..... 2076
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
137 IuAlaGluProPheAsnGluHisArgAspGlyLeuAlaAspLeuLeu 153
2077 GTGGTTTCCCGCAATGTGGCAAGTGAAGCGCATTTGGCATTTAAACAA 2126
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
154 PheValSerSerGlyProThrAsnAlaSerAlaPheThrGluArgAspAs 170
2127 TCAGCGCCCAACGACATTTTCGGTGTGCGACCGCATCAAGC..... 2170
: : : : : : : : : : : : : : : : : : : : : : : : :
170 nProSerGlnAspSerLysThrGlyMetLeuProCysAspSerPheAlaSer 187
2171 CAATCTGATACCGTTGGAGCTGAGC.....GCTGACACAGTGT 2211
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
187 hTrAlaValValSerGlnGluThrSerValGluAlaProAsnSerProCys 203
2212 ACCGAAAAAACCATTTACCGACGATMAAGTATGCTTCATTTAGACAGAC 2261
: : : : : : : : : : : : : : : : : : : : : : : : :
204 SerGlnSerCysValSerProGluValThrIleGluThrLeu...Gln.P 219
2262 CGACATCAGAGCGCAATGTACACCTTGGCATCACGCTCATTTAAATCTCA 2311
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
219 roAlaThrGluLeuSerLysAlaLeuAlaGluValGluSer.....ValLys 233
2312 CAGACTTGGCCACACCTCACAGCGCAATCTAGTGGAGGCGGACACGCCAC 2361
: : : : : : : : : : : : : : : : : : : : : : : : :
234 GluGlnLeuPro...AlaLysAlaLeuGluThrMetAlaGluGlnThrTh 249
2362 TATACGGTTACGGCGCA.....ACGCCACCCAAACGGCAACCT 2399
| : : : : : : : : : : : : : : : : : : : : : : : :
249 RasPvalValHisSerProSerThrAspThrThrProGlyProAspThrG 266
2400 CAGCTCTGTGGGCAATGCCACAGCAACATTTAATCAAGCCACA..... 2442
: : : : : : : : : : : : : : : : : : : : : : : : :
266 IuAlaAlaLeuAlaLysAspIleGluGluThrLysProAspValIle 282
2443 TTAAGCGCAACACATCGGCTTCGACACATGCTTCATTTAATGAGCAA 2492
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
283 LeuAlaAsnValThrGlnProSerThrGluSerAspMetPheLeuAlaG1 299
2493 CAACGCCGTACAAAACGCGAGCTGACGCTTTCGACACACGCTAAGGCAA 2542
: : : : : : : : : : : : : : : : : : : : : : : : :
299 nAspMet.....GluLeuLeuThrGlyThrGluAlaAlaHisAlaAla 313
2543 ACCTAACCATTCGCGCACTACAGCGCAATGTCTCCCTA.....GCC 2583
| : : : : : : : : : : : : : : : : : : : : : : : :
313 sn.....AsnIleIleLeuProThrGluPro 321
2584 GATPAGCGCATTTCCATTTTGAACACGCGGCTTACCGGAAAAAATCAG 2633
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
322 AspGluSerSerThrLysAspValAlaProPromeGluGluGluLeuVal 338
2634 CGGCGGCAAGAAATACGCAATTAACATAAAGACAGCAATGAGCGCTGC 2683
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
338 lProGlyAsnAspThrThr...SerProLysGluThrGluThrThrLeuP 354
2684 CG.....TCGGGC 2691
| : : : : : : : : : : : : : : : : : : : : : : : :
354 roIleLysMetAspLeuAlaProProGluAspValLeuLeuThrLysGlu 370
2692 ACGGAATTA.....GGCAATTTAAACCTTGACACACCGCACCAT 2729
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
371 ThrGluLeuAlaProAlaLysGlyMetValSerLeuSerGluIleGluG1 387
2730 TPACTCAATTCGCGCTATTCGACAGACATGCGGACGCGGCAAAACGGCA 2779
: : : : : : : : : : : : : : : : : : : : : : : : :
387 uAlaLeu.....AlaLysAsnAspValArgSerAlaGluIleProV 401
2780 GTGGCGCAAGATGGCGCGCGCGCGCTTCGGCGGCTCCATATATCCGTT 2829
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
401 alAlaGlnGluThrVal.....ValSerGluThrGluValValLeuAla 415

```

```

2830 AGCGCCCACTTCGGCAGAAATCCGTTTCAACAGCTGACGTAAACGG 2879
    ||| : : : : : ||| : : : : :
416 ThrGluValValLeuProSerAspProIleThrIleThrLysAspVa 432
2880 CAAATTGAACGGTCAAGGCAATCCGTTTATGTGGAACCTCTTGCT 2929
    ||| : : : : :
432 ThrLeu.....ProLeuGluAlaG 439
2930 ACCGAGCGGCAATTTGAAGTGGGGAAGTTCCGAGGCACTAACCC 2979
    ||| : : : : : ||| : : : : :
439 IuArgProLeuValIThrAspMetIThrProSerLeuGluThrGluMetThr 455
2980 TTGGCTGTCAACATACCGGCAACGCCGTAACTCTGAGCAATTGAC 3029
    ||| : : : : : ||| : : : : :
456 Leu.....GlyLysGluThrAlaProProThrGluThrAs 467
3030 GGTAAGTGAAGAAAGACACACACCGCTGTCCGAAATCTTAATTCA 3079
    : : : : : ||| : : : : : ||| : : : : :
467 nLeuGluMetAlaLysAspMetSerProLeuProGluSer...GluValT 483
3080 CCTGCAAAACGACACGTCGATGGCGGCGCATGGCGTTATCAGCTTATC 3129
    ||| : : : : : ||| : : : : :
483 nLeuGluLysAspValValIleLeuProGlu..... 493
3130 CGCAAAAGCGGCGAGTTCGCTGCATATCCGGTCAAGAACAGACGCT 3179
    ||| : : : : : ||| : : : : :
494 ThrLysValAlaGluPheAsnValIThrProLeuSerGluGluValVa 510
3180 TTCGCAGC.....AACTGGCAGGCGGGAAGAACAGAGCGGCTTGA 3223
    : : : : : : : : : : : : : : : : : : : : :
510 IThrSerValLysAspMetSerProSerAlaGluThrGluAlaProLeu. 526
3224 CGCAAAACGACACACTTGGCCGCAAAACAGAGCGGGAAGAACAC 3273
    ||| : : : : : ||| : : : : :
527 ..AlaLysAsnAlaLysPheIleHisSerGlyThrGluLeuIleValAsp 542
3274 GCGCAAGC.....CTTGACGCGCTGATTCG 3299
    : : : : : ||| : : : : :
543 SerMetAlaProAlaSerAspLeuAlaLeuProLeuGluThrLysValAl 559
3300 GGGC.....GGGCGCATGGCCGCAAAAGCGCAAGAGTGTG 3337
    ||| : : : : : ||| : : : : :
559 aThrValProIleLysAspLysGlyThrValGlnThrGluGluLysPro 576
3338 CGAACCAGCGCGGCGAGCGGGAAGTCCGCGC..... 3375
    ||| : : : : : ||| : : : : :
576 rGluAspSerGlnLeuAlaSerMetGlnHisLysGlyLysSerThrVal 592
3376 .....ATTATGCAAGCGGAGAGAGAGAAAA 3401
    : : : : : ||| : : : : :
593 ProProCysThrAlaSerProGluProValLysAlaAlaGluGlnMetSe 609
3402 AGGGGTGCAAGCGGATTAAGACACCGCTTGGCAAGACGCGGAGGCG 3451
    : : : : : ||| : : : : :
609 rThrLeuProIleAspAlaProSerProLeuGluAsnLeuGluGlnLysG 626
3452 AAACCGCGCGGCTACACCGCTTCCCGCGCGCGCGC..... 3490
    ||| : : : : : ||| : : : : :
626 IuThrProGlySer.GlnProSerGluProCysSerGlyValSerArgG 642
3491 .....GGGCGCGCGGATTTGGCGGCAAC.....CGCA 3518
    ||| : : : : : ||| : : : : :
642 nGluGluAlaLysAlaAlaValGlyValIThrGluLysAsnAspIleThrP 659
3519 GCGCCAAACGCAACCGCAACGCGAGCGGCACTGATCAGCGTTATGCCA 3568
    ||| : : : : : ||| : : : : :
659 rProLysnLysGluProProProSerProGluLysLysAla...LysPro 674
3569 ATAGCGTTTGAATTTTCCGCCAGCTCAACAGCGTTTTCGCGCTA 3618
    : : : : : ||| : : : : : ||| : : : : :
675 LeuAla.....ThrThrGlnProAlaLysThrSerThrSerLysAl 688
3619 CAGGACGAATTGGACCGGCTGTTGGCGAAGACGCGCGCA..... 3658

```

```

688 aLysThrGlnProThrSer...LeuProLysGlnProAlaProThrThrS 704
3659 .....ACGCGTTTGGACACAGCGCATCCGGGACACCA 3691
    ||| : : : : : ||| : : : : :
704 eArgGlyLeuAsnLysLysProMetSerLeuAlaSerGlySerValPro 720
3692 AACACTACCGTTGCGAAGATTTCCGCGCTACCGCCAAC..... 3730
    : : : : : ||| : : : : : ||| : : : : :
721 AlaAlaProHisLysArgProAlaAlaAlaThrAlaThrAlaArgProse 737
3731 .....AACCG...ACCTGGCCAAAT 3749
    ||| : : : : : ||| : : : : :
737 rThrLeuProAlaArgAspValLysProLysProIleThrGluAlaLysV 754
3750 CGGTATGCAGAAAACCTCGCAGCGGCGCGCTGCGATCCGTTTTCGC 3799
    : : : : : ||| : : : : : ||| : : : : :
754 aAlaGluLysArgThrSerProSerLysProSerSerAlaProAlaLeu 770
3800 ACAACCGAGCGGCAACACCTTGACAGCAGCGCATCGCAACTCGGACGG 3849
    ||| : : : : : ||| : : : : :
771 LysProGlyProLysThrThrProThrValSerLysAlaThr..... 784
3850 CTTGCCACGCGTCCGTTTTCGGGCAATACGCGATCGCAGTTCCACAT 3899
    ||| : : : : : ||| : : : : :
785 .....SerProSerThrLeuValSerThrGlyProS 795
3900 CGGCATCAGCGCGGCGGCTTTTATGACGCGAGCCTTTCAGACGGCA 3949
    ||| : : : : : ||| : : : : :
795 eSerArgSerProAlaThrThrLeuProLysArgProThrSerIleLys 811
3950 TCAGAGGCAAAATCCGCGCGCGCTGCTGCATTAAGGATTCAGGCAAGA 3999
    ||| : : : : : ||| : : : : :
812 ThrGluGlyLysProAlaAspValLysArgMetThrAla..... 824
4000 TACCGCGCAGTTTCCGCGGAT...TCGCGATGAAACCGCACATCGGCGC 4046
    : : : : : ||| : : : : : ||| : : : : :
825 ...LysSerAlaSerAlaAspLeuSerArgSerLysThrThrSerAla. 839
4047 AACCGCTATTGTGTCCAAAAAGCGGATTAACGATACGAAAGATCATA 4096
    ||| : : : : : ||| : : : : :
840 .....SerSerValLysArgAsnThr...ThrProThrGlyAla 851
4097 TCGCCACCGCGGCGCTTGATTCACCGCTTACGCGCGCGCATTAAGCA 4146
    ||| : : : : : ||| : : : : :
852 AlaProProAlaGlyMetThrSerThrArgValLys..... 863
4147 GATTATTCATTCAAAACGCGGCAACATTTCCATCAGCGCTTATTGAG 4196
    : : : : : ||| : : : : :
863 ..... 863
4197 CTTGCTCTATACCGATCCGCTTCCGCAAAAGTCCGAACCGCG...TCA 4243
    ||| : : : : : ||| : : : : :
864 .....PrometSerAlaProSerArgSerSerGlyAlaLeuSer 876
4244 ATACCGCGGATTTGCGCGAGGATTTGCGCAAAACCGCGACTCGCGAATG 4293
    : : : : : ||| : : : : : ||| : : : : :
877 ValAspLys.....LysProThrSerThrLysProSer..... 887
4294 GCGCTAAACGCGCAATCAAGGTTTCAGCGCTGCTCCGCTCAGCGTGGCG 4343
    : : : : : ||| : : : : : ||| : : : : :
888 .....SerSerAlaProArgValSerArgLeuAlaThrThrValSerA 902
4344 CGCCA.....AGGGCGCGCAATTTGAAGCGACGACGACGCG 4378
    ||| : : : : : ||| : : : : :
902 IaProAspLeuLysSerValArgSerLys.ValGlySerThrGluAsnI 918
4379 CGGGCATCAA 4388
    ||| : : : : : ||| : : : : :
918 eLysHisGln 921

```

seq_name: SwissProt_40:TCNA_TRYCR


```

1311 GGGCGTTATATACGTATGGCAGTACCGTACTTGAAGTAACGGCG 1360
      |||
274 .....LeuTrpLeuThrAsp... 278
      |||
1361 TGGCAAGACGCGCTGTCCAAATCGCAAAAGGACCGCTGCTGTCAA 1410
      |||
279 .....AsnGlnArgIle..... 282
1411 GCCAAGGGGAAACCAAGGCTCGGTACGCGTGGCGAGCGTAAGTCAT 1460
      |||
283 .....TyrAsnValGlyGlnValSerIleGlyAsp..... 292
1461 CTTAGATACAGCGCGGACGATCAAGCAAAACACGCTTTTACTGAAA 1510
      |||
293 .....GluAsnSerAlaTyrSerSerV 300
1511 TCGGCTGTGTCAGCGGCGAGGGGAGCGGTGCAACTGATCCGATATAG 1560
      |||
300 al.....Leu 301
1561 TTCAACCCCGACAACTCTATTTCGGCTTTCGCGGAGCGATTGGATT 1610
      |||
302 TyrLysAspAspLysLeuTyr..... 308
1611 GAACGGGCAATTCGCTTTCGCTTCCACCGCATTCATAATCCGATGA... 1656
      |||
309 .....CysLeuHisGluIle...AsnThrAspGluValT 319
1657 .....GGGCGATGATTCACACGC..... 1677
319 yTrSerLeuValPheAlaArgLeuValGlyGlnLeuArgIleLysSer 335
1678 .....AATCAAGACAAGAAATCCACGTTTACCATTTACAGCAATA...A 1718
      |||
336 ValLeuArgSerTrpLysAsnTrpThrAlaThr.CysProAlaPheAlaP 352
      |||
1719 AGATTTACTACACCGCGCAATTAACAACAATTGGATAGCAAAAAAGAAA 1768
      |||
352 rGlnLeuLeuIleGlnProLeuArgArgGlnArgValValValProLeu 368
1769 TTGCGTACACAGGTTGGTTGGCG.....AGAAGATGCAACCAACG 1812
      |||
369 SerProAlaGlyLeuValLeuLeuAlaPheCysArgGlnArgLeuProLysAr 385
      |||
1813 AACGGGCGGCTCAATCTGATATACCAACCGGAAAGCGGATCGCATTT 1862
      |||
385 gMetGlyGlySerTyrArgCysValAsn..... 394
1863 ACTGCTTTCGGCGGACAAATTTAAACGGCAATATCAAGCAACAAACG 1912
      |||
395 .....AlaSerThr 397
1913 GCAACGTGTTTTTCAAGCGGACAGCGACCGCGCT...ACAATCAT 1959
      |||
398 AlaAsn.....AlaGlnArgValArgAsnGlyLeuLysPheAl 410
1960 TTAGGAAGCGGCTGTCAAAAATGGAAGTATCCCAAGAGAAATCGT 2009
      |||
410 agLysValGlyGlyAlaLeuTrpProValSerGlnGlnGlyAsnG 427
2010 GTGGGACAACGATTTGGATCGACCGACAT...TTAAGCGGAAAACCTTC 2056
      |||
427 InaArgTyrArgPheAlaAsnHisAlaPheThrLeuValAlaSerValThr 443
2057 ATATTGAGGCGGACAGCGGTGTTTCCCGCAATGTT.GCCAAAGTGA 2105
      |||
444 IleHisGlnAlaProArgAlaAlaSerProLeuLeuGlyAlaSerLeuAs 460
2106 AGGCGATTGGCATTTAAGCAATACGCCCAACGATTTCGTGTGCGAC 2155
      |||
460 p.....SerSerGlyGlyLysLysLeuLeuGlyLeuSerT 472
2156 CGCATCAAAAGCCACACATCTGTACACGTTGGACTGGAGGGGTGTGACA 2205
      |||
472 yAspGlnLysHis.....GlnTrpGlnProIleTyr 482
2206 AGTTGTACGGAATAAACCATTACCGAAGATTAAGTGATGTTCTCATAG 2255
      |||
483 GlySerThrProValThrProThr..... 490
2256 CAAGACCGACATCAGAGGCAATGTCAAGCTTCCGATCAACGCTATTAA 2305
      |||
491 .....GlySerTrpGlnGlyLysArgTyrHisLeuVal 502
2306 ATCTCACA.....GACTTGCACACTCAACGCAATCTT 2340
      |||
502 alLeuThrMetAlaAsnLysIleGlySerValTyrIleAspGlyGlnLeu 518
2341 AGTGCAGCGGAGACACACGCACTATACGGTTACGCGACGCCACCAAAA 2390
      |||
519 LeuGlnGly..... 521
2391 CGGCACCTCAGCCTCGTGGCAATGCCCAAGCAACATTTAATCAACCA 2440
      |||
521 ..... 521
2441 CATTAACGGCAACACATCGGCTTTCGACAAATGCTTCATTATATACG 2490
      |||
522 .....SerGlyGlnThrValAlaProAspLysArgThrProAspIleSer 536
2491 AACACCGCTTACAAAACGCGAGCTTGACGCTTTCGCAACGCTAAGGC 2540
      |||
537 HisPheTyrVal.....GlyGlyTyrLysArgSerAsp.....MetTr 549
2541 AACGTAAAGCAATTCGCGACTCAACGCAATGTCTCCCTAACCGGATAGG 2590
      |||
549 oThrIleSerHisValThrValAsn...AsnValLeuLeuTyrAsnArg 565
2591 CAGATTTCATTTGAAAACAGCGGCTTTACCGGAAAATACAGCGCGCG 2640
      |||
565 rGlnLeuAsnThrGlnLeuIleArg..... 573
2641 AAGGATACGCAATTTACCTTAAAGACAGCAATGAGCGTGGCGGCGG 2690
      |||
574 .....ThrLeuPheLeuSerGlnAsp.....LeuLeuG 583
2691 CACGGAATTAAGCAATTTAAACCTTGACAACGCCACCATTTACACTCAAT 2740
      |||
583 yThrGln.....AlaHisMetAspSerSerAspSerAlaHis 597
2741 CCGGCTATCGACACGATGCGGCGAGCGCGCAACCGGACAGTGGGAGAT 2790
      |||
597 lSerThrProSerThrProAlaAspSerSerAlaHisSerThrProSer 613
2791 GCGCGCGCGCGCTTTCGCGCGCTCCCTATTTACGCTTACGCGGCAAC 2840
      |||
614 ThrProValAspSerSerAlaHisSer.....ThrProSerThr 626
2841 TTGGGACGAA...TCCCGTTTCAACAGCTGACGAGTAACGCAAAATGA 2887
      |||
626 rProAlaAspSerSerAlaHisGlyThrProSerThrProValAspSerS 643
2888 ACGGTACGGAACATTCGCTTATGTCGGAACCTTTCGGCTACCGGACG 2937
      |||
643 eAlaHisGlyThrProSerThrProAlaAspSerSerAlaHisGlyThr 659
2938 GGCMAATGAAGCTGGCGGAAGTTCCGAAGCACTTACACTTGGCTGT 2987
      |||
660 ProSerThrProValAspSerSerAlaHisSerThrProSerThrProVa 676
2988 CAAACATACCGGACAGCAACCGCTAAGCTTCGAGCAATTCAGCGTGTGG 3037
      |||
676 lAspSerSerAlaHisSerThrProSerThrProValAspSerSerAlaHis 693
3038 AAGGAAAAGACACACGCGCTGTGCGAAATGTTAATTTCACCGCTGCAA 3087
      |||

```



```

693 IsGlyAlaProSerThrProAlaAspSerSerAlaHisGlyThr.Prose 709
3088 AACGACACGTCGATGCGCGCATGCGTTACGTTATCGCGCAAAA 3137
709 rThrProValAspSerSerAlaHisGlyThrProSerThrProAlaAspS 726
3138 CGCGGAGTTCCGCGTCATATCCGGTCAAGAACAAAGACTTTCGACA 3187
726 eSerAlaHisSer.....ThrProSerThrProAla 736
3188 AACTCGGCAAGGCGGAGAAACAGAGCCGCTTACGCGCAAAACAGGCA 3237
737 AspSerSerAlaHisSerThrProSerThrProAlaAspSerSerAlaH 753
3238 CAATCGCGCG.....CCAAACACAGCGCGGAAAAAGACACAGCGCGAAG 3281
753 sSerThrProSerThrProValAspSerSerAlaHisGlyThr..... 767
3282 CCTGACGCGCTGATTGCGCGCGGCAATGCCACCGGAAAGGAGANA 3331
768 .....ProSerThrProAlaAspSerSerAlaHis 777
3332 GTGTTGCGCAACCGCGCGCGAGCGCGGAAAAATCGCGCATTTATG 3381
778 SerThrProSerThrProAla..... 784
3382 CAGGCGGAGAGAAAGAAAAACGGGTGACGCGGATTAAGACACCGCGCTT 3431
785 .....AspSerSerAlaHisGlyThrProSerThrProVal 796
3432 GCGGAAACAGCGGAAAGCGGAAACCGCGCGCTACACCGCGCTTCCGCC 3481
796 AlaAspSerSerAlaHisSerThrProSer...ThrProValAspSerSer 811
3482 GCG.....CCGCGCGCGCGCGCGGATTTGCGCGAACCGGACGCC 3522
812 AlaHisGlyThrProSerThrProValAspSerSerAlaHisSerThrPr 828
3523 CAACCGCAACCCCAACCGCGCGCGCATGACGCGCTTATCGCAATAG 3572
828 oSerThrProValAspSerSerAlaHisGlyThrProSerThrProVal. 844
3573 CGGTTGAGTGAATTTCCGCCACGCTCAACAGCGTTTCCGCGGTACAG 3622
845 .....AspSerSerAlaHisSerThrProSer 853
3623 ACGAATTGACCGCGGTGTTGCCGAAGCGCGCGACGCGCTTGAGCA 3672
854 ThrProAlaAspSerSerAlaHisSerThrProSerThrProAlaAspSe 870
3673 ACGCGCATCGGAGACACAAACACTACCGTTGCAAGATTTCGCGGCTA 3722
870 rSerAlaHisGlyThrProSerThrProValAspSerSerAlaHisSer 887
3723 CGGCAACAAACCGACCTCGCCAAATCGGTGCAAGAAAACTCGGCA 3772
887 hrProSerThrPro.....AlaAspSerSerAlaHisSerThrProSer 901
3773 GCGGCGCGCTCGCATCTGTTTGCACA..... 3802
902 ThrProValAspSerSerAlaHisSerThrProSerThrProAlaAspSe 918
3803 .ACCGACCGGAAACACTTCGACAGCGCATGCGCACTCGCACGCGCT 3851
918 rSerAlaHisGlyThrProSerThrProValAspSerSerAlaHisGlyT 935
3852 TGCCACGCGTGGCGTTTGGCGCATACGCGCATCGCAGGTTGACATCG 3901
935 hrProSerThrProAlaAspSerSerAlaHisSerThrProSerThrPro 951
3902 GCATACGCGCGCGCGGTTTAGTAGCGGCGACGCTTCAACAGCGCATC 3951
952 AlaAspSerSerAlaHis.....SerThrProSerThrProAlaAspSe 966

```

```

3952 AGAGCAAAATCCGCGCGCGGTGTCGATTACGCGATTCAGGCAATGA 4001
966 rSerAlaHisSerThrProSerThrProValAspSerSerAlaHisSer 983
4002 CGCGCAGGTTTCGCGGATTCGCGCATGCAACCCACATCGCGCAACGC 4051
983 hrProSerThrProAlaAspSerSerAlaHisSerThrProSerThrPro 999
4052 GCTATTGTCGCAAAAGCGATTACCGATACGAAACGTCATATCGCC 4101
1000 AlaAspSerSerAlaHisSerThrProSerThrProAlaAspSerSer. 1015
4102 ACCCGCGCGCTTCGATTCAACGCGTACCGCGCGGCGCATTAAGGAGATTA 4151
1016 .....AlaHisSerThrProSerThrPro..... 1024
4152 TTCATTCAACCGCGCGCAACATTTCCATCAGCGCTTATTTGAGCGCTG 4201
1024 AlaAspSerSerAlaHisSerThr...ProSer..... 1033
4202 CCTATACCGATGCGCGCTTCGCGCAAGTCGACAGCGCGGTCAATACGCC 4251
1034 .....ThrProAlaAspSerSerAlaHisGlyThrPro.. 1044
4252 GTATTGCGCGAGATTTCGCAAAACCGCGACGTGCGGATGGGCGTAAA 4301
1045 .....SerThrProAlaAspSerSerAlaHisSer 1055
4302 CGCGCAAAATCAAGGTTTCACGCTGTCCTCCACGCTGCGG 4342
1055 hrProSerThrProValAspSerSerAlaHisSerThrPro 1068

```

seq_name: SwissProt_40:P2P_LACPA

```

seq_documentation_block:
ID P2P_LACPA STANDARD; PRT; 1902 AA.
AC 002470;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE PII-type proteinase precursor (EC 3.4.21.96) (Lactocopein) (Cell wall-
DE associated serine proteinase) (LP151).
GN PRP.
OS Lactobacillus paracasei.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;
OC Lactobacillus.
OX NCBI_TaxID=1597;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NCDO 151.
RX MEDLINE=92381481; PubMed=1512565;
RA Holck A., Naes H.;
RT "Cloning, sequencing and expression of the gene encoding the cell-
RT envelope-associated proteinase from Lactobacillus paracasei subsp.
RT paracasei NCDO 151."
RT J. Gen. Microbiol. 138:1353-1364(1992).
RN [2]
RP SEQUENCE OF 189-196.
RX MEDLINE=9222694; PubMed=1564442;
RA Naes H., Nissen-Meyer J.;
RT "Purification and N-terminal amino acid sequence determination of the
RT cell-wall-bound proteinase from Lactobacillus paracasei subsp.
RT paracasei".
RL J. Gen. Microbiol. 138:313-318(1992).
RL
CC -1- FUNCTION: PROTEASE WHICH BREAKS DOWN MILK PROTEINS DURING THE
CC GROWTH OF THE BACTERIA ON MILK.
CC CATALYTIC ACTIVITY: Endopeptidase activity with very broad
CC specificity, although some substrate preference have been noted,
CC e.g. large hydrophobic residues in the P1 and P4 positions, and
CC pro in the P2 position. Best known for its action on caseins,
CC although it has been shown to hydrolyse hemoglobin and oxidized
CC insulin B-chain.

```



```

995 helysleuaspserlysalaprotthrvalarghisvalalauserala 1011
1108 ACAGCAGAGAACCTTTATCATCTGCAGGTGGGTCAACAGTTATCG 1157
1012 LysThr..... 1013
1158 ACCCAGACTGATTAATGGAGAAATATTTCTTTATGCAGAAAGAAAG 1207
1014 .....GluAsnGlyLysT 1018
1208 GTGATTTGATCTTACACACATCAACCAAGCGCGCGTGTGTAT 1257
1018 hrcInrtyrleuthralaglualalaspserleu... 1033
1258 TTTGAGGTAATTTACGGTCTCGCTTAAACAAACGAAACGTGGCAAG 1307
1034 .....AspAlaThrLysSerValLysThr..... 1041
1308 CGGGGCGTTCAATCAATGATGGCAGTACCGTTACTTGAAAGTAACG 1357
1042 .....AlaIleAsnG 1045
1358 GCGTGGCAACGACCGCTGTCCAAATCGCAAGGACGCTGCTGT 1407
1045 LvalThrAsn.....Leu 1049
1408 CAAGCCAAAGGGAACCAAGCGCTGTCAGCGTGGCGAGGTAAAT 1457
1050 ..... 1066
1050 AspAlaThrPheThrAspAlaGlyThrThrAlaAspGlyTyrThrLysI 1066
1458 C.....AQTATGATCAGCAGCGGACGATCAAGCAAAACAAAG 1498
1066 egluthrProleuSerAspIuInlaInlaInlaInlaInlaInlaInla 1083
1499 CCTTTAGTAAATCGGCTGTGTCAGCGGAGGCGGAGGTCAATGAT 1548
1083 snserAlaGlyLeuTyrLeuThrAsp.....Asn 1092
1549 GCGGATATACGTTCAACCCCGCAACCTTATTC.....GGCTTTCG 1592
1093 AlaSerAsnAlaThrAsnGlnAspAlaSerValGlnLysProGlnSer 1109
1593 CGGCGGACGTTGATTTGAACGGCATTCGCTTTCCTCCACCGCATTC 1642
1109 rserPheAspLeuIleValAsnGlyGlyIlePro...AspLysIleS 1125
1643 AAAATACCGATGAAGGGCGATGTTGTCAACCAATCAAGCAAGAA 1692
1125 erSerThrThrThrGly.....TyrGlnAlaAsnThrGlnGly 1138
1693 TCCACCGCTTACCATACAGCAAT.....AAGATTTTCTAC 1730
1139 GLyThrTyrThrPheSerGlyThrTyrProAlaAlaValAspLysThr 1155
1731 AAC.....GGCAATAACAACAACCTTGATAGCAAAAGAAATG 1771
1155 rThAspAlaGlnGlyLysLysHisAspLeuAsnThrThrTyrAspAla 1172
1772 CCTACACGGTGTGGCGAGAAAGATGCAACCAACGAAAGCGGCG 1821
1172 laThrAsnSerPheThrAlaSerMetAlaValThrAsnAlaAspTyrAla 1188
1822 CTCATCTGAATTCACACCGGAAGAGCGGATCGCACTTACTGCTTC 1871
1189 AlaGlnValAspLeuTyrAlaAspLysAlaHisThrGlnLeuLeuShi 1205
1872 CGGGGACAACAATTTAAACGGCAATATCAGCAACAAACGCAACTGT 1921
1205 sPheAspThrLysVal..... 1210
1922 TTTTACGGGACGACACCGCGCTACATCTTTAGCAAGCGG 1971
1211 .....ArgLeuThrAlaProThrPheThrAspLeuLysPhe... 1222
1972 TGTCAAAATGGAAGTATCCACAGAGAAATGCTGTGGACACAGCA 2021
1223 .....AsnAs 1224
2022 TTTGATGACCGCATTTTAAAGCGGAAACCTTCATATTTCAGCGCGAC 2071
1224 nGlySerAspIleThrSerGlnAla...ThrIleLysValThrGlyThrV 1240
2072 AAGCGGTG.....GTTCCCGCAATGTTGCCAAAGTG 2103
1240 alSerSerAspThrLysThrValAsnValGlyAspThrValAlaAlaLeu 1256
2104 GAAGGCGATGCGATTATAGCAATCAGCCCAACGATTTTCGGTGTGCG 2153
1257 AspAlaGlnHisHisPheSerValAspValProValAsnTyrGly... 1271
2154 ACCGCATCAAGCCACACAAATC...TGTACAGTTCGAGCTGACGCGGT 2200
1272 .....AspAsnThrIleLysValThrAlaThrAspLysPglY 1285
2201 TGACAGTTGTACGGAACCAATTCACCGACGATTAAGTATGCTTCA 2250
1285 snThrThrThrGlnGlnLysThrIleThr.....Ser 1295
2251 TTGACAGACAGCCATCAGACGATGTCACCTTCGCCATCAGCT... 2298
1296 SerTyrAspProAspValLeuLysAsnAlaValThrPheAspGlnGly 1312
2299 .....CATTTAATCTC... 2310
1312 LysPheGlyAlaAsnGlnPheAsnAlaThrSerAlaLysPheTyrAsp 1329
2311 .....ACAGGACTTCGCACACTCAACGCG... 2334
1339 rOlysthrGlyLeuAlaThrIleThrGlyLysValLysHisProThrThr 1345
2334 ..... 2334
1346 ThrLeuGlnValAspGlyLysGlnIleSerIleLysAsnAspLeuThrPh 1362
2335 .....AATCTTACTGCAGGC..... 2349
1362 eSerPheThrLeuAspLeuGlyThrLeuGlyGlnLysProPheGlyVal 1379
2350 .....GGAGACAGCATATACGTTACGCGGACCCACCAACGCGC 2394
1379 alValGlyAspThr.....ThrGlnAsnLysThrPheGlnGly 1391
2395 AACCTCAGCGCTGTGGCAATGCCACAGCAACATTTAATCAAGCCACAT 2444
1392 AlaLeuThrPheIleLeuAspAlaValAlaPro...ThrLeuSerIle 1406
2445 AAAGGCAACACATCGGCT.....TGAGACATGCTTCATTTAATC 2485
1406 uAspSerSerThrAspAlaProValTyrThrAsnAspProAsnPheGlnI 1423
2486 TAAAGCAACAGCGGTACAA.....AAC 2508
1423 leThrGlyThrAlaThrAspAsnAlaGlnTyrLeuSerLeuAlaIleAsn 1439
2509 GGCAGTCTGACGCTTTCGACACAGCTAAGGCAACGTAAGC..... 2550
1440 GlySerHisValAlaSerGlnTyrAlaAspIleAsnIleAsnSerGlyLys 1456
2551 .....CATTCGCACTGACGCAATGCTTCCTTACCGAT..... 2586
1456 sProGlnHisMetAlaIleAspIleProValLysLeuLeuGlnGlyLys 1473
2587 .....AAGCAGTATTCATTTTGAACACGCGCTTACCGCA 2625
1473 snValLeuThrValAlaValThrAspSerGluAsnAsnThrThrLys 1489
```

2626 AAATCAGCGCGGCAAGATACGGCATTTACACTTAAAAAGACAGGAATG 2675
 |||||::: :::: ||||
 1490 LysIleThrValTyrGluProLysIleuAlaIleProThVa 1506
 2676 GACGCTCGCGGCGGCAAGATTAAGCAATTTAACTTGACACAGCCA 2725
 ||||| ||||| ::::: |||||
 1506 Thr...ProSerThrThrGluProAlaLys...:::ThrValT 1518
 2726 CCATTACACTCAATTCGGCTTATCGACACGATGCGGCGGCGCAACC 2775
 ||::: ||||| |||||::: |||||
 1518 ThrThrAlaAsnAlaIle...:::AlaThrGlyLnuThrVal 1530
 2776 GCGAGTGGCGGCAATGCGCGCGCGCTTGGCGCGCTCCCTATATC 2825
 ::::: |||||::: ::::: |||||
 1531 GluTyrSerAlaAspGlyGlyLysThrTyrGlnAspValProAlaIle 1547
 2826 CGTTACGCGCGCACTTGCGGCAAGATCCGTTTCAACAGCTGACGTA 2875
 :||| :||| :||| :||| :||| :|||
 1547 Val...:::ThrValThrAla 1553
 2876 ACGCCAATTAAGCGTACGAGGCAATTCGGTTATGTC...GAATC 2922
 || ||||| |||||::: |||||
 1553 sn...:::GlyThrPheLysPheLysSerThrAspLeu 1563
 2923 TTGGGCTACCGGCAAGCAATTAAGCTGCGGCAAGCTCCGAGGAC 2972
 :||| :||| :||| :||| :||| :|||
 1564 TyrGly...:::AsnGluSerProAlaVal 1571
 2973 TTACACCTTGCTGACACATACCGGCAAGCAAGCTGATCGAGC 3022
 ::::: |||||::: |||||::: |||||
 1571 AspTyrValValThrAsnIleLysAlaAspProAlaGlnLeuGln 1587
 3023 AATGACGCTAGTGAAGAAAGAACACACACCGCTGCCGAAATCTT 3072
 ||||| :||| :||| :||| :||| :|||
 1588 ThrAlaLysGlnAlaLeuThrAsnLeuIleAlaSer... 1599
 3073 AATTACACCTGCAAAACGACACGCTGATGCCGCGCATGCTTATCA 3122
 1599
 3123 GCTTATCCGCAAGACGGCGAGTCCGCTGCATATCCGCTCAAGAC 3172
 1600
 3173 AAGAGCTTCCGCAAACTCGGCAAGCGGAGAAACAGAGCGCGCTTG 3222
 ::::: ||||| ||||| ||||| |||||
 1603 euseAlaSerGlyLysTyrAspAspAlaThrThrAlaLeuAlaIle 1619
 3223 ACGCCAAACGCGCAACTTGGCGGCAACACAGCGGCAAAAGACAA 3272
 :||| :||| :||| :||| :||| :|||
 1620 AlaThrGlnLysAlaGlnThrAlaLeuAspGln...:::Th 1631
 3273 CGCGCAAGCTTGACGCGCTGATTCGCGCGGCGGCAATGCCACGAAA 3322
 ||||| :||| :||| :||| :||| :|||
 1631 rAspAlaSerValAspSerLeuThrGlyAlaAsnArgAspLeuGlnThra 1648
 3323 AGGCAAGAAAGTGTTCGCGCAACCGCGCGGCAAGCGGCGGAAAAATGCC 3372
 :||| :||| :||| :||| :||| :|||
 1648 ThrAlaGlnLeuAla... 1653
 3373 GGCATTATGACGGGCAAGAAAGAAAGAGGCTGACGCGGATTAAGA 3422
 1654
 3423 CACCGCTTGGCGAAGACGCGGCAAGCAAGCAAGCGCGGCTACACCG 3472
 ||||| :||| :||| :||| :||| :|||
 1661 sThrSerLeuAsnGlnLeuInserValLysAlaIleGlyThr 1677
 3473 CATTCCCGCGCGCGCGCGCGCGCGGATTCGCGCAACCGACGCC 3522
 ||||| :||| :||| :||| :||| :|||
 1678 AspLeu...GlyAsnGlnThr... 1683

3523 CACCGCAACCCACCGGCGGCGGACCTGATACCGCTTATGCCAATAG 3572
 ::::: |||||::: ::::: |||||
 1684 AspProSerThrGlyLysThr... 1690
 3573 CGGTTGAGTGAATTTCCCGCACGCTCAACAGCGTTTCCGCGTACAG 3622
 1690
 3623 ACGAATTTGACCGCGTGTTCGCGAAGACCGCGCAACCGCGTTTGAGA 3672
 ::::: |||||::: |||||
 1691 PheThrAlaIleLeuAspAspLeuValAla...G 1701
 3673 AGCGGCATCCGGGACACCAACACTACCGTTGCAAGATTTCGCGCTTA 3722
 ||||| :||| :||| :||| :||| :|||
 1701 AlaGlnAlaGlyThrGlnThrAlaAspGlnLeuGlnAlaSerLeuAla 1718
 3723 CCGCAACAAACCGACCTCGCGCAAAATCGGTATGCAAAACCTTGCGA 3772
 ::::: |||||::: |||||
 1718 ValLeuAspAlaValAlaLeuAlaLys...LeuAla 1728
 3773 GCGGCGCGCTCGCATCTGTTTTCGACACCGGCAACCGCAACCTTC 3822
 ||||| :||| :||| :||| :||| :|||
 1729 GluGlyIleLysAlaIle...ThrProAlaGluValGlyLys 1741
 3823 GACGAGCGCATCGCACTCGGACGCGCTTGGCCACGAGTTCGCTTTCG 3872
 :||| :||| :||| :||| :||| :|||
 1741 AlaLysAspAlaIleThr...GlyLysThrTyrAlaAspIle 1756
 3873 GCAATACGGCATCGGCGAGTTCGACATCGCATACGCGGCGGCGGTT 3922
 ::::: |||||::: |||||::: |||||
 1756 LaspThrLeuThrSerGlyGlnAlaSerAlaAspAlaSerAspLysLeu 1772
 3923 TTGATGCGCGGCACTTCAGACGCGCATCAGAGGCAAAATCCGCGCGC 3972
 ||||| :||| :||| :||| :||| :|||
 1773 AlaHisLeuGlnAlaLeuGlnSerLeuLysThrLysValAla 1786
 3973 GTGCTGATTAAGCATTCAGGCA... 3997
 1786 AlaAlaValAlaGlnAlaAlaLysThrAlaGlyLysGlyAspThrThrng 1803
 3997
 1803 LysThrSerAspLysGlyGlyGlyGlyThrProAlaProAlaProGly 1819
 3998 GATACCGCGGATTTCCGGGATTCGGGATTCGGGATTCGGGATTCGGG 4047
 |||||::: |||||::: |||||::: |||||
 1820 AspThrGlyLysAspLysGlyLysGlySerGlnProSerSerGlyG 1836
 4048 ACGCCGATTTGCTCCAAAACCGGATTAACGATACGAAACGTCATAT 4097
 :||| :||| :||| :||| :||| :|||
 1836 LysAlaLeuProThrLysProAlaThrThrThrSerThrSerThrAsp 1853
 4098 CGCGACCGCGGCGCTTGATTAACCGCTACCGCGCGGCGGCTTAAGG 4147
 :||| :||| :||| :||| :||| :|||
 1853 ThrThrAspThrAsnGlyGlnHisThrSerGlyLysGlyAlaLeuProLys 1869
 4148 ATTATTCATTAACCGCGGCAACATTT 4177
 :||| :||| :||| :||| :||| :|||
 1870 ThrAlaGluThrThrGluArgProAlaPhe 1879

seq_name: SwissProt_40:STFR_ECOLI
 seq_documentation_block:
 ID STFR_ECOLI STANDARD; PRT; 1120 AA.
 AC P76072; P7560;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Side tail fiber protein homolog from lambdaoid prophage Rac.
 GN STFR OR B1372.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia.

NCBI_TaxID=562;
 [1]
 RN SEQUENCE FROM N.A.
 RC STRAIN-K12 / MG1655;
 RX MEDLINE=97426617; PubMed=9278503;
 RA Blacther F.R., Plunkett G., IIT, Bloch C.A., Perna N.T., Burland V.,
 RA Riley J.M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RT "The complete genome sequence of *Escherichia coli* K-12.";
 RL Science 277:1453-1474(1997).
 RN
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12;
 RX MEDLINE=97251357; PubMed=9097039;
 RA Alpha H., Baba T., Fujita K., Hayashi K., Inada T., Isono K., Itoh T.,
 RA Kasai H., Kashimoto K., Kimura S., Kitakawa M., Kitagawa K.,
 RA Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,
 RA Nakade S., Nakamura Y., Nishimoto H., Nishio Y., Oshima T., Saito N.,
 RA Sempel G., Seki Y., Sivasubraman S., Tagami H., Takeda J.,
 RA Takemoto K., Takeuchi Y., Wada C., Yamamoto Y., Horiuchi T.;
 RT "A 570-kb DNA sequence of the *Escherichia coli* K-12 genome
 corresponding to the 28.0-40.1 min region on the linkage map.";
 RL DNA Res. 3:363-377(1996).
 CC -1 SIMILARITY: BELONGS TO THE TAIL FIBER FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: AE000234; AAC74454.1; ALT_INIT.
 DR EMBL: D90774; BAI14966.1; -;
 DR EMBL: D90775; BAI14975.1; -;
 DR HSSP: P04002; IMFA.
 DR EcGene: EG13370; stfr.
 DR InterPro: IPR004089; Chemotaxis_transducer.
 DR Hypothetical protein; Fiber protein; Repeat; Complete proteome.
 KW SEQUENCE 1120 AA; 113779 MW; 542E59D71EB795B4 CRC64;
 SQ

alignment_scores:
 Quality: 196.50 Length: 530
 Ratio: 0.685 Gaps: 19
 Percent Similarity: 54.151 Percent Identity: 23.208

alignment_block:

US-09-303-518D-653 x STFR_ECOLI ..

Align seg 1/1 to: STFR_ECOLI from: 1 to: 1120

```

2968 GGCACCTTACACCTTGGCTGTCAACATACCGGCAAGAACCGGTAGTCT 3017
      ||| |||:||||: ||| |||
56 GlyTyrSerValIleLeuLeuValGluGlyPheProProSerHisAl 72
3018 CGAGCAATTCAGCGTAGTGGAGAGAAAGACACACCGCTGTCCGAA 3067
      :|||:||||: ||| |||
72 aglyThrIleThrValTyrGluAspSerGlnProGlyThrLeuAsnAsp 89
3068 ATCTTAATTTACACCTTGGCTGTCAACATACCGGCAAGAACCGGTAGTCT 3117
      |||:||||: ||| |||
89 helGluLeuMetValGluGluValAlaArgAsnAlaSerAlaValAlaG 105
3118 TATCAGCTTATC.....CGCAAGACGGCGAGTTCGGCCT 3152
      :|||:||||: ||| |||
106 PheGluLeuMetValGluGluValAlaArgAsnAlaSerAlaValAlaG 122
3153 GCATTAATTCGGCTCAAGAACAGCTTCCGCAACACCTCGCAAGCGG 3202
      :|||:||||: ||| |||
122 nasThrAlaAlaLysSerAlaSerAlaSerThrSerAla 139

```

```

3203 GAGAA.....ACAGGCGCGCTTGACGGCAAAACAGGCA 3237
      |||:||||: ||| |||
139 rgGluAlaIleThrHisAlaAlaLysPheAlaLysSerAlaValAla 155
3238 CAATTCGCGCAACACAGCGGAGAAAGACACCGCAAGCCTTGA 3287
      :|||:||||: ||| |||
156 SerThrSerGluGlyGlnAlaAlaSerSerAlaGlnSerAlaSer 172
3288 CGCGCTGATTCGGCGCGGCGCAATGCACCGAA.....AAGCAGAA 3331
      |||:||||: ||| |||
172 rAlaGlyThrAlaSerThrLys...AlaThrGluAlaSerLysSerAla 188
3332 GTGTTCGCAACCGCGCGGCGGAGCGGAGAAATGCCGATTATG 3381
      :|||:||||: ||| |||
188 lAlaIleAlaGluSerSerLysSerAlaAlaIleThrSerAlaGlyAla 204
3382 CAGCGCGAGAGAGAGAA.....AAGCGGTGCAGCGGCA 3416
      :|||:||||: ||| |||
205 LysThrSerGluThrAsnAlaSerAlaSerLeuGlnSerAlaIleThr 221
3417 TAAAGACACCGCTTGGGAGAAACAGCGCAAGCGGAAC.....CGG 3460
      :|||:||||: ||| |||
221 rAlaSerThrAlaThrThrThrLysAlaSerGluAlaIleThrSerAla 238
3461 CGGCTACACCGCGCTTCCCGCGCGCGCGCGCGCGCGGATTGCG 3510
      |||:||||: ||| |||
238 sPheAlaAlaIleSerLysGluAlaLysSerSerGluThrAsn..... 252
3511 CAACCGCAGCCCAACCGCAACCCCAACCGCAGCGCACTGATCACC 3560
      |||:||||: ||| |||
253 .....AlaSer 255
3561 TTATGCCATAGCGGTTGAGTGAATTTCCGCCAGCTCAACAGCCTT 3610
      |||:||||: ||| |||
255 rSerAlaSerAlaIleSerAlaIleThrAlaIleGlyAsnSerAla 272
3611 TCGCCGTACAGAGCAATTTGAGC...GCCGTTTTCGCAAGACCGCG 3657
      |||:||||: ||| |||
272 yAlaAlaLysThrSerGluThrAsnAlaArgSerSerGluThrAla 288
3658 AACGCCGTTTGACAGCGGATCCGGGACCAACACTACCTTGCA 3707
      |||:||||: ||| |||
288 acGlyIleSerAlaSerAlaIleAlaGlySerLysThrAlaAlaLys 305
3708 AGATTTCGCG.....CCTACCGCAACAAACCGACCTGCCAATTC 3751
      :|||:||||: ||| |||
305 eAlaSerAlaIleSerThrSerAlaGlyGlnAlaSerAlaSerAla 321
3752 GTATGCAGAAACCTCGGACGCGCGCGCTGCGCATCTGTTTCGCAC 3801
      :|||:||||: ||| |||
322 AlaAlaGlyLysSerAlaGluSerAlaIleAlaSerSerAlaThr 338
3802 AACCGACCGCAACACCTTGCAGACGCGCATGCGCACTGCGACGCT 3851
      |||:||||: ||| |||
338 rThrLysAlaGlyGluAlaThrGluGlnAlaSerAlaAlaLysSer 355
3852 TGCCACGCGTCCGTTTCGGGCAATACGCGATCGCGGATTCGACAT 3901
      :|||:||||: ||| |||
355 lAserAlaLysThrSerGluThrAsnAlaLysAlaSerGluThrSer 371
3902 GCATCAGCGCGCGCGGCTTTAGTACGCGGACCTTTCAGACGATC 3951
      |||:||||: ||| |||
372 AlaGluSerSerLysThrAlaAlaLysSerAlaSerAlaLys 388
3952 AGAGCAAAATCCGCGCGCGCGCTGTCATACGCACTTCAGGCA...G 3998
      :|||:||||: ||| |||
388 rSerAlaSerSerAlaSerAlaSerLysAspAlaIleThrArgGlnAla 405
3999 ATACGCGCAGGTTTCGGGATTCGCGCATCAACCG..... 4036
      :|||:||||: ||| |||
405 eAlaAlaLysSerSerAlaThrThrAlaSerThrLysAlaThrGluAla 421

```

```
4037 .....ACATC 4041
422 AlaGlySerAlaThrAlaAlaIaGlnSerLysSerThrAlaGlnSerAl 438
4042 GGGGCAACGGCGTATTTCGCCAAAAGCGGATACCGATACGAACGT 4091
438 aaIaThrArgAlaGlnThrAlaAlaIaLysArgAlaGlnAspIleAlaSerA 455
4092 CAATATCGCCACCCGGCCCTGCATTCAACCGCTACCGCGGCGCATTA 4141
455 laVal.....AlaLeuGlnAspAlaSerThrThrLys..... 465
4142 AGCGCATTTATTCATTCAAAACCGGCGCAACACATTTCCATCAGCCTTAT 4191
466 LysGlyIleValGlnLeuSerSerAlaThrAsnSerThrSerGluThrIle 482
4192 TTGAGCGCTGCTCTATACCGATGCGGCTCCGGCAAGTCCGAACGGCGT 4241
482 uAlaIaIaThrPro.....LysAlaValLysSerA 492
4242 CAATACCGCGCGTATTGGCGCAGGATTCGGCAAAACCGCAGTGGCAAT 4291
492 laTyTrAspAsnAlaGlnLysArgLeuGlnLysAsp.....GlnAsn 505
4292 GGGGCGTAACCCGCAA.....TCAAAGT 4317
506 GlyAlaAspIleProAspLysGlyCysPheLeuAsnIleAsnAlaVa 522
4318 TTCACGCTGCTCCCTCCACGCTGCGCGCAAGGGCGCGCAATTGGAAGC 4367
522 lSerLys.....ThrAspPheAlaAspLysArgGlyMetArgIyValA 537
4368 GCAGCACAGCGGCGCATCAAAATTAGCTACCGCTGT 4405
537 rgValAsnAlaProAlaGly.....AlaThrSerGly 547
```